Computers for Chemistry and Chemistry for Computers: From computational prediction of reaction selectivities to novel molecular wires for electrical devices

Severin Thomas Schneebeli

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

Abstract

Computers for Chemistry and Chemistry for Computers: From computational prediction of reaction selectivities to novel molecular wires for electrical devices

Severin Thomas Schneebeli

Taking advantage of cutting-edge technologies in computational and experimental chemistry, my Ph. D. research aimed to bridge both of these chemical subdivisions. Therefore, while part I of this dissertation focuses on new structure-based computational methodologies to predict selectivities of organic and enzymatic reactions, part II is concerned with the design, the synthesis and the electrical properties of novel, single molecular wires. These single molecule technologies described in part II are likely to contribute to more powerful computer chips in the future, which will in turn lead to faster and more accurate computational predictions for chemical problems.

Part I: Computers for Chemistry: Progress towards the design of accurate computational tools to predict the selectivity of chemical reactions

The first fully quantum mechanical study to predict enantioselectivities for a large dataset of organic reactions has been reported. Enantioselectivities were calculated for a diverse set of 46 dioxirane catalyzed epoxidation reactions. Comparison to experiments showed that our methodology is able to accurately predict the free energy differences between transition states leading to enantiomeric products. To further improve the predictive performance, we have also developed a new correction scheme, which increases the accuracy of density functional theory (DFT) for non-covalent interactions. Our new correction scheme accurately estimates interaction energies of non-covalent complexes not only with large, but also with small basis sets at lower computational cost. The improved enantioselectivity prediction protocol containing our latest non-covalent

corrections has now been fully automated in a user-friendly fashion. We are currently testing its accuracy for other asymmetric reactions, such as CBS reductions and are also trying to use our methodology to design new asymmetric organocatalysts. In collaboration with Dr. Jianing Li, a structure based computational methodology to predict sites of metabolism of organic substrates with P450 enzymes has also been developed, which is highly relevant for structure-based drug discovery.

Part II: Chemistry for Computers: From novel antiaromatic and π-stacked molecular wires to highly conducting link groups with direct Au-C bonds

Part II of this dissertation describes studies of antiaromatic and π-stacked molecular wires as well as new direct ways to connect them to gold electrodes. At the beginning, the first successful single molecule conductance measurements ever on partially antiaromatic molecular wires are described. These wires, based on a biphenylene backbone, were synthesized via a highly regioselective cyclization enabled by the antiaromaticity. Then, two new ways to connect single molecules to gold electrodes with direct Au-C links are presented. The first methodology is based on strained arene rings in [2.2]-paracyclophanes, which were found to directly contact gold electrodes with their π-systems. The second methodology employs tin based precursors, which get replaced in situ by gold electrodes to also form direct Au-C bonds with very low resistance. The direct Au-C bonds observed with strained paracyclophanes enabled us to study, for the first time, single molecule conductance through multiple layers of stacked benzene rings. Further single molecule conductance studies with less strained stacked benzene rings are

currently under way and will provide additional valuable evidence about electron transport in stacked π-systems.

Table of Contents

List of Abbreviations

CYP	Cytochrome P450
LJ	Lennard-Jones
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
OPLS	Optimized Potentials for Liquid Simulations
OPLS-AA	Optimized Potentials for Liquid Simulations-All Atom
PBE	Poisson Boltzmann Equation
PDB	Protein Data Bank
PLOP	Protein Local Optimization Program
QM	Quantum Mechanics
RMSD	Root-Mean-Square Deviation
SOM	Sites of Metabolism
NSEC	Nanoscale Science and Engineering Center
CCI	Center for Molecular Optimization of Electronic Plastics
UB3LYP	Unrestricted B3LYP
MUE	Mean Unsigned Error
BSSE	Basis Set Superposition Error
DFT	Density Functional Theory
BEGDB	Benchmark Energy & Geometry Database
SCF	Self Consistent Field Method
QM/MM	Mixed Quantum Mechanics molecular Mechanics

CPU	Central Processing Unit
STM	Scanning Tunneling Microscope
HOMO	Highest Occupied Molecular Orbital
LUMO	Lowest Unoccupied Molecular Orbital
CV	Cyclic Voltammetry
TLC	Thin Layer Chromatography
HPLC	High-performance Liquid Chromatography
NMR	Nuclear Magnetic Resonance
MS	Mass Spectrometry
CI	Chemical Ionization
FAB	Fast Atom Bombardment
MALDI	Matrix assisted Laser Desorption Ionization
THF	Tetrahydrofurane
DMF	Dimethylformamide
DCM	Dichloromethane
PPA	Polyphosphoric acid
DCE	Dichloroethane
TBAP	Tetrabutylammonium perchlorate
HF	Hartree Fock Theory
RHF	unrestricted Hartree Fock Theory
DMSO	Dimethylsulfoxide
NBS	*N*-Bromosuccinimide

List of Figures

Part I. Computers for Chemistry

Part II. Chemistry for Computers

List of Schemes

Part I. Computers for Chemistry

Part II. Chemistry for Computers

List of Tables

Part I. Computers for Chemistry

Acknowledgements

I am incredibly grateful to Professor Ronald Breslow and Professor Richard Friesner, who both supported and guided me as my advisors during my five years of Ph. D. Research at Columbia University. They have inspired me with research ideas and have granted me adequate freedom to explore a variety of scientific fields.

Special thanks also goes to all other members of my graduate research committee, Professor Colin Nuckolls, Professor David Reichman and Professor Philip Kim, who have given me helpful advice for my graduate research and this dissertation.

I would also like to thank all the current and previous members of the Breslow and Friesner groups, who collaborated with me on research projects, shared their valuable research experience with me and were also always available to discuss my research ideas. Special thanks go to Dr. Shengping Cheng, Dr. Jordan Quinn, Dr. Frank Foss and Dr. Craig Kenetsky, who broadened my view about the art of synthesis at the beginning of my Ph. D. research. I would also like to especially acknowledge Dr. David Rinaldo and Dr. Arteum Bochevarov, who introduced me to a lot of concepts and details important to computational chemistry and with both of whom I had the great pleasure to discuss some of my early research ideas as well. Special thanks also goes to all my collaborators (if not already mentioned above) in the Friesner and Breslow research groups, Michelle Lynn Hall, Dr. Jianing Li, Dr. Thomas Hughes, Andrew Weisman, Dr. Rachid Skouta, Dr. Zhanling Cheng, Dr. Wenbo Chen, Alice Chang and Christine Yeah.

As illustrated in this thesis, I have also been collaborating with several other research groups during the past five years at Columbia, mostly as part of the Nanoscale Science and Engineering Center (NSEC), but during my fifth year as part of the Center

for Molecular Optimization of Electronic Plastics (CCI) as well. I am very grateful to all of my collaborators, especially however to Professor Latha Venkataraman and her graduate students Maria Kamenetska, Jonathan Widawsky and Brian Capozi. I would also like to thank Professor Katherine Willets, Dr. Mark Hybertsen, Dr. Hector Vazquez, Dr. Namdong Kim, Dr. Leif Sherrill, Sarah Stranahan, Dr. Kyle Plunkett, Dr. Kateri DuBay for their collaborations and for helpful discussions.

Finally, I would like to thank my fiancée, Dr. Jianing Li, for her never-ending support and advice during the last five years. She has also been a great collaborator and helped me to proof read this dissertation.

In the end, I am also very grateful to the Guthikonda Family and to Pfizer, *Inc.*, for supporting part of my graduate research with an Arun Guthikonda Memorial Fellowship and a Columbia-Upjohn Fellowship.

This dissertation is dedicated to
my Father, Mother, Fiancée, and all those,
whose support, encouragement, and personal sacrifice
have made this research possible

Preface

I was very fortunate to work on a multitude of projects at the interface of computational and experimental chemistry during the five-year studies towards my Ph. D. degree with Professor Ronald Breslow and Professor Richard Friesner. I worked on computational methods development, but also had great opportunities to explore the synthesis and the properties of molecules with potential applications for nanotechnology. This interesting, but often challenging mix of projects allowed me to gain experience in a lot of different fields centered around chemistry from programming and computational data analysis all the way to organic synthesis. The fundamental idea of all my research projects was to combine and integrate computational and experimental chemistry, as it is my strong belief that computational chemistry will be used one day by experimental chemists on a routine basis as a powerful predictive tool. As a result it will not just help saving lives by allowing for faster development of better drugs, but it might also save our environment by allowing us to develop affordable alternative energy sources. Therefore I put my efforts in trying to advance the predictive power of computational chemistry and also used chemical approaches to design and study the behavior of novel molecular wires, which might become parts of faster computers in the future.

My dissertation focuses on two examples of successfully integrating computational and experimental organic chemistry, which are described in two separate parts: Part I focuses on the use of computers to predict chemical reactivity, while part II describes the discovery of novel single molecular wires, knowledge of which might help to build powerful molecular computers in the future.

Part I. Computers for Chemistry: Progress towards the design of accurate computational tools to predict the selectivity of chemical reactions

1 Introduction

1.1 Introductory remarks about computational predictions

The increasing power of computers, which has become available at constantly decreasing cost, has made computational methods an invaluable tool for chemists to understand and design new complex chemical systems. In order to show how the importance of computational chemistry has increased over the past decade just in the field of organic chemistry, I plotted the percentage of *J. Org. Chem.* and *Org. Letters* publications, which include computational investigations, in Figure 1.1. The figure shows a clear trend of an increasing fraction of organic papers that involve computational investigations.

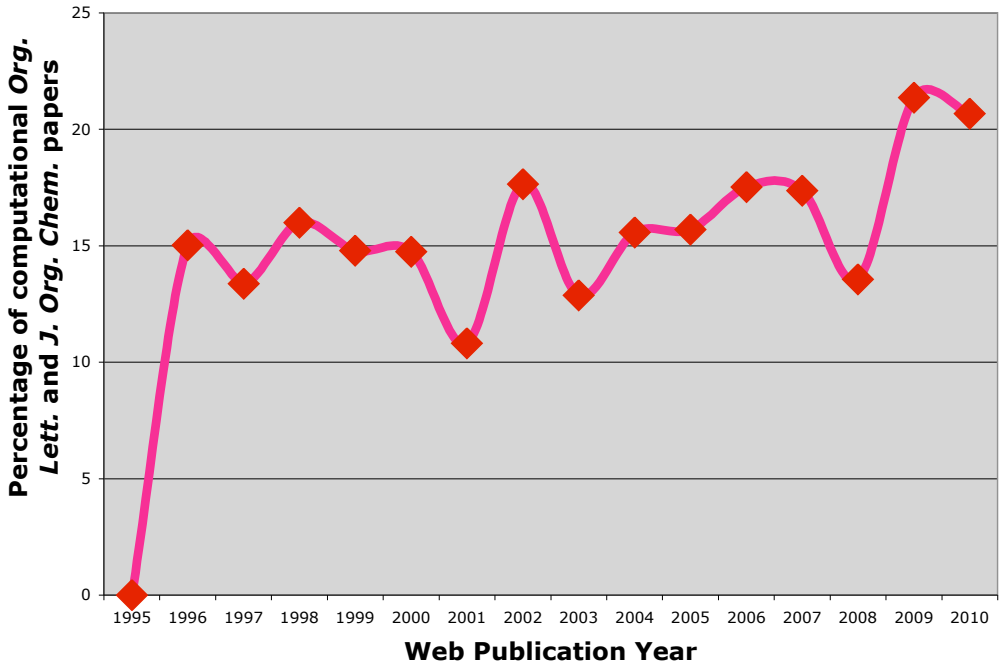


Figure 1.1. The percentage of publications containing computational investigations in organic chemistry journals has been increasing steadily over the past decade. The percentage of such publications was estimated by counting the number of hits returned by performing a search in the *Journal of Organic Chemistry* and *Organic Letters* for the term "computational*" on the ACS website for all years shown.

One great reason for using computational methods in chemistry is that they can in principle provide an understanding of chemical processes at a level of detail which is often not available experimentally. Transition state modeling provides one such example. While transition states and reaction paths can be studied computationally almost as simply as ground states, experimental information about these unstable points on the energy surface can often only be obtained indirectly, for example by measuring the activation parameters of reactions. Since the lifetimes of transition states are only on the order of a chemical vibration, spectroscopy with a femtosecond time resolution is needed in order to study transition states directly experimentally. The Zewail research group has pioneered the field of femtosecond laser spectroscopy, which was recognized by the 1999 Nobel prize in chemistry and to the first direct experimental study of transition state structures.[1,2] However, even though transition states can be detected with superfast detection technologies, detailed structural information is still hard to obtain with such experiments. In addition, most of these superfast detection technologies require expensive equipment and are far from trivial to carry out. On the contrary, computational methods are becoming increasingly powerful at a quickly decreasing cost and have also become easier to use with the development of user-friendly graphical interfaces,[3] which help to set up the calculations. Recently, many chemistry educators have also started to include some computational coursework in their curriculum, which will help to educate more chemists about computational methods. This new generation of chemists, which "grew up" with computational methods will likely feel more comfortable using computational tools on a routine basis for their daily research than the current generation, such that computation might finally become part of almost all high quality research.

Apart from the structural and energetic details provided by calculations, computational approaches have the additional advantage that properties of any structure can be predicted before the molecule has ever been synthesized. Therefore, computational predictions are able to help us prioritize the spending of time and money on the synthesis of molecules with desirable properties, greatly accelerating new research discoveries in chemistry. One of the most successful examples is virtual screening, which has become a routine technique for many pharmaceutical companies to speed up the drug discovery process.[4]

In summary, computational predictions have great potential to streamline the discovery of new chemical solutions to many important problems, such as healthcare and sustainable development. However, there are still a lot of challenges waiting to be solved to make computational predictions even more helpful for chemical discovery. In the following section, these challenges will be discussed in more detail.

1.2 What is needed to make successful computational predictions

As introduced above, computational tools ideally enable us to reveal all the details about a chemical system, such as structures and energies of all the conformations. However, one major difficulty, which arises for the computational study of many chemical systems, is the enormous conformational space growing exponentially with increasing molecule size in general. This often makes it impossible to study all low energy conformations of a certain state with an accurate computational method, even for medium sized organic molecules. A popular solution to this problem is to use hierarchical methods to handle the searching of conformational space. In such a hierarchical approach initial conformational sampling is carried out with low-cost computational methods. The

structures and energies of only the lowest energy conformations are then further refined with a more accurate method.

While hierarchical computational approaches are in principle straightforward to implement, they quickly become very labor intensive if not carried out automatically with computer scripts, simply because there are usually a very large number of conformations to process. Since only a few fully automatic workflows are already implemented in an easy to use fashion in the currently available software packages, this requires the users to learn a scripting language and become familiar with details of the software packages, such as input and output file formats (which unfortunately often vary between different software packages).

For computational investigations of transition states an additional practical problem arises. While conformational searching for ground state structures can relatively easily be carried out with molecular force fields, most current force fields fail to correctly describe transition states. Even though a lot of effort has been invested in trying to build generally applicable transition state force fields, such approaches often only work well for transition states of specific reactions and should therefore be used with great caution for other systems.[5] As an alternative to molecular mechanics force fields, semiempirical molecular orbital methods can be used to carry out a limited amount of conformational searching of transition state structures as well.[6] However, since such semiempirical approaches are also highly parametrized, one has to also use them very carefully, especially if applied to reactions, which were not in the training set used to fit the parameters of the method.[7] Furthermore, it is also important to note that transition state search algorithms can easily fail to converge structures to the desired transition states,

especially if started from a poor initial guess. Therefore, optimization of multiple low energy transition states of even a simple organic reaction often involves a lot of trial and error and is by no means a simple procedure.

All of the above mentioned difficulties limit the widespread use of computational predictive tools among experimental chemists, who usually are not familiar with all these details. Therefore, it is very important that computational chemists simplify the currently available computational methodologies, such that experimentalists will be able to use them without needing to learn all the underlying details. Such a goal can for example be achieved by developing automatic workflows to predict important problems, which should be easily accessible via graphical interfaces.

Apart from the bottleneck in exploring the large conformational space, another reason limiting the practical use of computational predictions is the well-known trade-off between computational cost and accuracy. Even though nowadays correlated wave-function based QM methods (such as coupled cluster) are available in many commercial software packages providing accuracy sufficient for most chemical problems,[8] these methods are too expensive to be run on larger systems of chemical interest. Therefore, one has to resort to less accurate methods such as DFT, which treat electron correlation only approximately and therefore give results of variable accuracy, depending on the systems studied.[9]

The most crucial part of a DFT based approach is the density functional, which describes approximately how electron correlation affects the total energy of a molecule.[10] The most successful DFT functionals available today all contain a significant amount of empirical character and therefore the accuracy of these functionals is system-dependent.

While it is clearly desirable to develop new density functionals with as little empiricism as possible, a small amount of empirical parameters seems to be necessary in order to achieve reasonably accurate descriptions of molecules. A good balance between accuracy and empiricism is provided by the B3LYP[11-14] functional,[15,16] which has therefore become the most widely used functional for chemical problems. While B3LYP displays unsatisfactory accuracy for certain problems, such as for the description of dispersion interactions,[17-20] Professor R. Friesner and coworkers have shown that these errors are often systematic and can therefore be corrected with incremental corrections containing only few empirical parameters.[21-25]

Other researchers such as Professor D. Truhlar and coworkers have also been developing new density functionals. Their initial benchmarking work has shown improved accuracy over B3LYP with some of the functionals developed in their group, which represents recent progress in this area.[26-28] However, as most of these functionals contain a relatively large number of empirical parameters in combination with complex functional forms,[27] there are worries about overfitting and the generality of these newer functionals. As described later in this thesis, Michelle Hall and I applied both B3LYP and M06-2X, one of the most promising new functionals from the Truhlar group, to a large dataset of dioxirane epoxidations and found that B3LYP clearly outperforms the M06-2X functional for these systems.[29] This result further raises concerns about the generality of Professor Truhlar's new density functionals, especially when considering that dioxirane epoxidations were not part of the training set used by the Truhlar group to parameterize the M06-2X functional.

Therefore, while DFT currently seems to provide the best performance which balances computational cost and accuracy, careful benchmarking studies with large and diverse datasets are still in need to confirm its accuracy for problems of general chemical interests. Such benchmarking studies might reveal systematic inaccuracies of DFT to varying extent, which can then be fixed by adding additional empirical correction terms, as has been demonstrated in this dissertation for the B3LYP functional and non-covalent interactions.[21]

In summary, I believe that the often unpredictable accuracy of cost effective computational methods as well as the specialized knowledge required to set up calculations of complicated chemical systems remain some of the most important barriers, which prevent the widespread use of computational methods by experimental chemists. Therefore, the next section will introduce some of my Ph. D. work, the goal of which has been to try to overcome some of the aforementioned barriers for the general use of computational predictions in chemistry.

1.3 Predicting selectivities of organocatalyzed and enzymatic reactions

In order to overcome some of the aforementioned major barriers, which are still preventing widespread use of predictive computational chemistry, I decided to develop new computational methodologies with predictive power for practical problems in reactive organic and enzymatic chemistry. One important practical problem in chemistry is the development of selective catalysts for enantioselective synthesis.[30] In the past, the selectivity of such catalysts has mostly been improved by a combination of experimental trial and error and intuition of organic chemists.[31,32] As experimental trial and error

approaches are generally very time intensive, we thought that a more general, structure based methodology to predict enantioselectivity would be of high value to the community of synthetic chemists and started our first project towards this goal at the beginning of my Ph. D. research.

Apart from being useful to speed up asymmetric syntheses, predicting enantioselectivity of chemical reactions also provides a valuable, but challenging test for the accuracy of the currently available computational methods, with a large amount of experimental data already available in the literature due to the increased interest in enantioselective catalysis in the near decade.

The main challenge in trying to accurately predict enantioselectivities comes from the fact that very small free energy differences between transition states on the order of ca. 0.5 kcal/mol already lead to relatively large changes of the observed enantioselectivities. This means that the accuracy of the computational methodology has to be at least as high as these small energy differences involved in significantly changing the enantioselectivities. What makes such predictions even more challenging is that most enantioselective reactions are carried out in a solvent, which can significantly affect the relative energies of the transitions states as well.

As described in detail in the following chapter, we therefore collected a diverse experimental dataset of 46 asymmetric dioxirane-catalyzed epoxidation reactions of olefins,[29] a reaction which finds widespread use in organic synthesis.[33] We were then able to show that density functional theory is indeed capable of correctly predicting the small relative transition state free energy differences involved in determining the enantioselectivity of these reactions. While we found that the B3LYP functional gave

surprisingly high relative accuracies with a mean unsigned error below 0.5 kcal/mol, we soon realized that the well known deficiencies of B3LYP in treating dispersion interactions[17-20] were likely responsible for some serious outliers in our dataset.

Therefore, in order to further improve the predictive power of our methodology, I and Dr. Arteum Bochevarov, a former postdoc in the Friesner group, then developed empirical corrections for B3LYP, which were shown to greatly improve non-covalent interaction energies on a very large dataset of thousands of benchmark non-covalent interaction energies.[21] While our new correction scheme will be described in detail in chapter 3, it worth to mention one key point, which distinguishes our correction scheme from similar ones devoped by others[34-36]. This key difference is that our correction scheme also works well with small basis sets of reduced computational cost. This is especially important, since the use of large basis sets for the transition state structure optimizations needed to predict enantioselectivities is simply too expensive and therefore not practical at the moment. We also showed that the commonly used approach of using a small basis set for the structure optimizations followed by single point energy calculations leads to small random energy fluctuations, which are large enough to significantly worsen the predictive accuracy of the enantioselectivity prediction methodology. Therefore, we believe that our correction scheme for non-covalent interactions will be especially valuable for predicting enantioselectivities since it also delivers accurate results with small basis sets of lower computational cost.

After successfully implementing our new correction scheme, I completely automated our prediction methodology for enantioselectivities in order to make it easy to use. I am currently also in the process of generalizing our methodology to new enantioselective

reactions, such as the well known Corey-Bakshi-Shibata (CBS) reduction[37]. Our newly developed corrections for non-covalent interactions, which should help to improve the predictive power of our methodology, are also being applied to our greatly increased datasets now containing multiple types of reactions. These more recent benchmark tests and improvements to our prediction protocol will be summarized in chapter 4.

While organic chemists nowadays make a lot of use of catalysts comprising of simple organic molecules, which sometimes also contain a transition metal or two, nature has developed much more sophisticated catalysts to carry out her chemical reactions. Nature uses enzymes, which contain thousands of atoms and often also transition metals. While the computational modeling of simple organic catalysts already represents a challenging problem due to all the reasons mentioned above, the accurate computational modeling of enzymatic systems is even more challenging, mainly due to the increased size and flexibility of the catalysts. A detailed understanding and prediction of enzymatic chemical reactivity is however of enormous importance for many problems relevant to chemistry, one of the most prominent ones being drug discovery. Enzymatic reactivity is important for drug discovery, since most drugs eventually get metabolized by enzymes and can get converted to intermediates with undesirable properties, such as high toxicity.[38] Since the P450 superfamily of enzymes is mainly responsible for the initial metabolism of drug compounds in the human body,[39] Dr. Jianing Li and I developed a novel, structure based methodology, which accurately predicts the sites of metabolism (SOM) for potential drugs with P450 enzymes, which is briefly summarized in Chapter 5.

Prediction of drug metabolites is very important for drug discovery, as it will help to direct the synthetic efforts of the experimentalists towards drug candidates with

desirable metabolic behavior and therefore help to speed up the drug discovery process.[40] Computational prediction of metabolic reactions catalyzed by P450 enzymes is a very challenging task, mainly due to the great flexibility of the P450 active sites, which often show strong induced fit effects. Other reasons, which make such predictions difficult, are the complicated multistep catalytic cycle of the P450 enzymes as well as the fact that the catalytic center includes a transition metal center (Fe), which has multiple low energy spin states.[41] While it is possible to calculate transition states corresponding to different sites of metabolism with relatively accurate QM/MM methods,[42] such approaches are quite slow and therefore sufficient conformational sampling becomes almost impossible.[43] This represents a serious problem especially for larger ligands of relevant size to drug discovery. Therefore, significantly faster force field based approaches are currently the only alternative for predicting sites of metabolisms, which can be applied to a large amount of ligands while still allowing for sufficient sampling.

However, as already mentioned earlier, it is difficult to model transition states with such methods. Furthermore, the accuracy of force field based energy models for proteins has long been a limiting factor of the predictive power of such methods until the new VSGB2.0 energy model[44] developed by Dr. Jianing Li and coworkers finally seems to have at least partially surmounted this accuracy barrier.[45] The very accurate performance of the VSGB2.0 energy model for protein structure predictions[44] gave us hope that with this energy model accurate predictions of P450 based SOMs might be possible as well. Therefore, Dr. Jianing Li and I developed the first fast *and* accurate structure based protocol for P450 SOM predictions based on the VSGB2.0 energy model and an accurate quantum mechanical reactivity model.[46] Despite all of the

aforementioned challenges, we were indeed able to obtain transition state like structures of relatively large and flexible organic substrates reacting with P450 enzymes with our fast, force field based approach. Thanks to the relatively low computational cost, our approach allows efficient sampling of the induced fit effects and the ligand conformations. Our new structure based methodology will briefly be summarized in chapter 5 of this dissertation and has been described in much greater detail in the dissertation of Dr. Jianing Li.[45]

2 Quantitative computational prediction of enantioselectivity for dioxirane catalyzed epoxidation reactions

2.1 Introduction

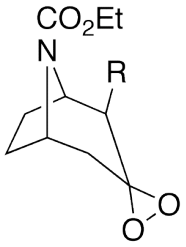
As already mentioned in the previous chapter, the development of tools to quantitatively model the stereochemical outcome of reactions is of great interest to the chemical community, especially since experimental screening of new compounds for enantioselective reactions is often labor-intense and expensive. Furthermore, modeling stereochemistry presents a formidable challenge to the computational chemist, where a difference of only 1.8 kcal/mol in the stereochemically distinct transition states leads to product ratios of 96:4.

Moitessier, *et al.* have recently proposed a computational tool to model diastereoselectivity employing a force-field based approach.[47,48] Moitessier and coworkers approximate transition state geometries as linear combinations of reactant and product geomtries while their "transition state energies" are calculated by multiplying the MM3* force field terms for the forming and breaking bonds with a linear scaling factor. Of course, such an approach has the advantage that the computations are orders of magnitude faster than quantum mechanical calculations, and therefore allow more efficient sampling of conformational space. However, the accuracy of such a crude transition state model is questionable, especially for flexible transition states. A quantum mechanical approach, however, promises higher accuracy for the transition state geometries and energies and does not require the parametrization of a transition state force field. The major disadvantage of a quantum mechanical approach is, of course, the much longer times required to run the calculations. However, the Jaguar[49] suite of quantum chemistry programs allows efficient computation of molecular integrals by

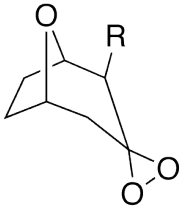
exploiting pseudospectral methods[50-54] and therefore opens the possibility to model enantioselectivity of reactions between medium sized organic molecules within affordable time. As a rough estimate, the computations described in this chapter took on the order of one week on a standard PC for a medium sized transition state in the dataset.

In this chapter the first fully quantum mechanical study of enantioselectivity for a large dataset is now reported. Our study differs significantly from previous quantum mechanical studies of asymmetric reactions, which have been largely mechanistic or limited to only a handful of substrates. Bach *et al.* first explored computational methods to help understand asymmetric epoxidations,[55-61] paving the way for Houk and coworkers[5,62-65] who later contributed much to the understanding of the reaction's stereoselectivity. Houk showed that stereoselectivity in asymmetric epoxidations is qualitatively controlled by torsional steering, a differential stability of transition states mediated by varying degrees of torsional strain.[5,62-65] It has also been shown that the dioxirane-mediated epoxidation, among other epoxidations, is well described by B3LYP.[5,62-69] Sarzi-Amadè and coworkers have published several studies exploring the qualitative nature of enantioselectivity for peroxy acids and dioxiranes, utilizing B3LYP to obtain good qualitative agreement with experimental data.[70-76] Singleton *et al.* have also demonstrated that B3LYP can be used to predict kinetic isotope effects in very good agreement with experiment for the Shi epoxidation,[68] further suggesting B3LYP's ability to accurately model dioxirane-mediated epoxidations even for larger complex systems. A recent study by Shankar and coworkers discusses the facial selectivity in the DMDO epoxidation of carbohydrate-based oxepines based on B3LYP calculations and compares the computational results to experimental findings.[77]

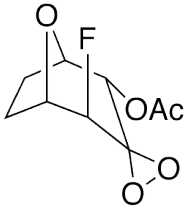
Intrigued by whether computational chemistry could be used to explain organic phenomena of practical interest quantitatively, Michelle Hall and I therefore employed transition state theory to model enantioselectivity on a large dataset. Our test set comprised 46 asymmetric dioxirane-catalyzed epoxidation reactions with dioxirane catalysts shown in Figure 2.1 and olefins shown in Figure 2.2. The dataset is notable for the structural diversity of the dioxirane catalysts and of the olefin substrates as well as the broad range of experimental ee-values (ranging from 0 to 98 % ee). Furthermore, the dataset is also marked by various solvent systems and reaction temperatures. By utilizing a dataset of such diversity, we aimed to thoroughly explore the boundaries and limitations of quantum mechanical methods to accurately reproduce experimental data with a keen eye on the future goal of prediction.



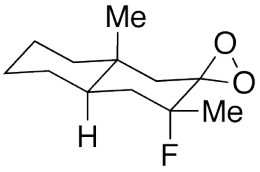
1: R = F
2: R = Cl
3: R = OAc



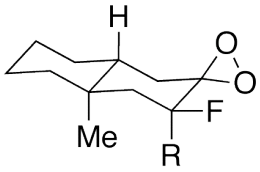
4: R = F
5: R = OAc



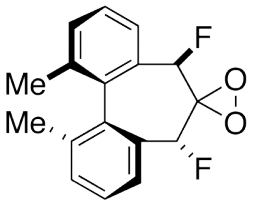
6



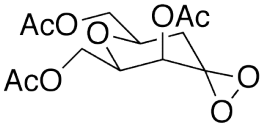
7



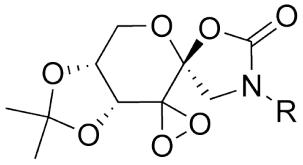
8: R = Me
9: R = H



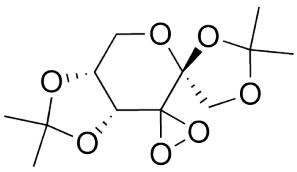
10



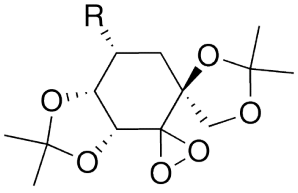
11



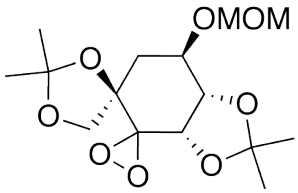
12: R = Boc
13: R = *p*-PhPh



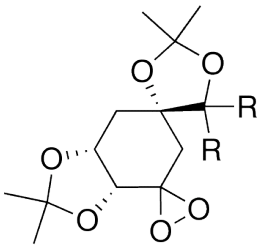
14



15: R = H
16: R = OMOM



17



18: R = H
19: R = Me

Figure 2.1. Dataset of dioxirane epoxidation catalysts.

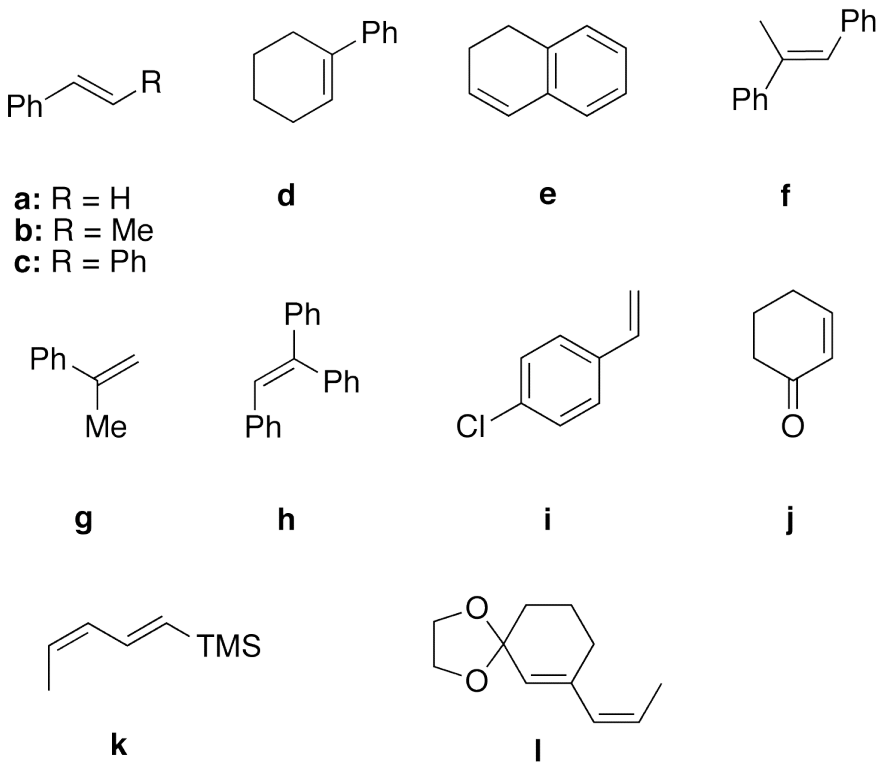
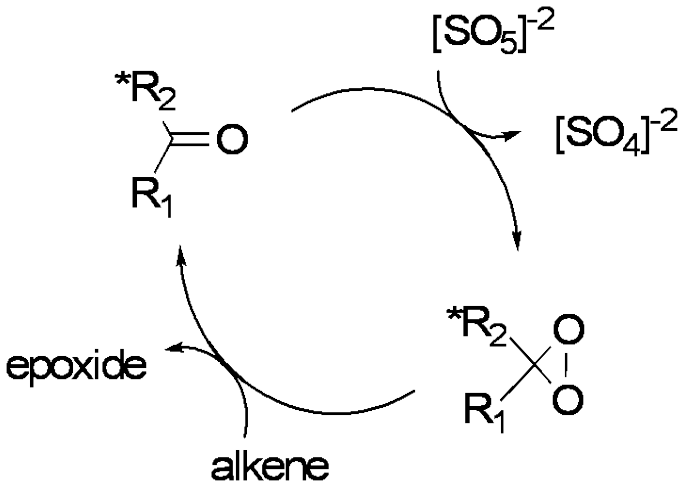


Figure 2.2. Dataset of olefin epoxidation starting materials.

Our current test set includes many practically useful dioxirane-catalysed epoxidations. The mechanism of the catalytic dioxirane-mediated epoxidation reaction of olefins is well understood and proceeds through the catalytic cycle shown in Scheme 2.1.[78] Computations by Bach *et al.*[55-61] and Houk *et al.*[63] have shown the epoxidation of olefins by dioxiranes to proceed through a "spiro" transition state as shown in Figure 2.3a. Although planar transition state structures (Figure 2.3b) are invoked by Shi some cases,[79-81] computational evidence continues to support a spiro transition state[61] or hybrid planar-spiro transition state in more complex systems, as shown by Singleton.[68]



Scheme 2.1. Catalytic cycle for the asymmetric epoxidation reaction of olefins with dioxiranes.

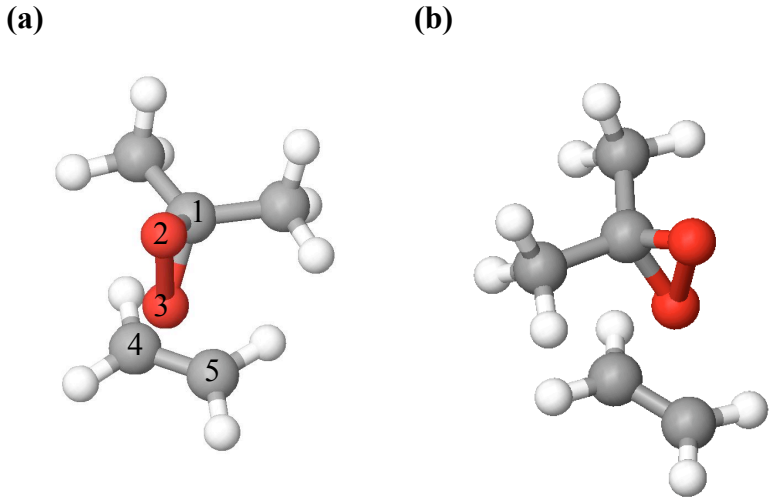


Figure 2.3. Spiro **(a)** and planar **(b)** transition states of dimethyldioxirane (DMDO) and ethylene.

2.2 Physical model and Computational Methodology

The enantiomeric excess is given by

$$ee = \frac{k_{favored} - k_{disfavored}}{k_{favored} + k_{disfavored}} \qquad (1.1)$$

where the k_i's are the reaction rate constants leading to either the favored or disfavored enantiomer. The rate constants can be expressed via transition state theory as

$$k = A(T)e^{-\Delta G^{\neq}/RT} \tag{1.2}$$

where $A(T)$ is a pre-exponential factor, assumed to be equivalent for both enantiomeric pathways, $\Delta G^{\neq}$ is the activation Gibbs free energy, R is the gas constant, and T is the temperature in Kelvin. Substitution of (1.2) into (1.1) then allows us to write

$$ee = \frac{\exp(-\delta\Delta G^{\neq}/RT) - 1}{\exp(-\delta\Delta G^{\neq}/RT) + 1} \tag{1.3}$$

Where $\delta G^{\neq}$ is given by

$$\delta\Delta G^{\neq} = \Delta G^{\neq}_{favored} - \Delta G^{\neq}_{disfavored} \tag{1.4}$$

As the starting materials leading to the two different enantiomers are identical, and therefore also have identical free energy, we can write the difference in Gibbs free activation energy, $\delta G^{\neq}$, just as the difference in transition state free energies, ΔG^{TS}.

$$\delta\Delta G^{\neq} = \Delta G^{TS} = G^{TS}_{favored} - G^{TS}_{disfavored} \tag{1.5}$$

Because the Gibbs free energy, G, is simply related to the Helmholtz free energy, A, as $G = A + pV$, where pV is assumed to be identical for both enantiomeric pathways, we set ΔG^{TS} equal to ΔA^{TS}. The Helmholtz free energy of a transition state ensemble *i*, was then computed from an estimate of the partition function, Z_i.

$$A_i^{TS} = -RT\ln Z_i \tag{1.6}$$

Neglecting entropic effects partially, the partition function Z_i was approximated as a Boltzmann-weighted average over unique transition states (shown schematically in Figure 2.4) leading to enantiomer *i*.

$$Z_i \approx \sum_j \exp\left(-\frac{E_j^{TS}}{RT}\right) \tag{1.7}$$

E_j^{TS} is the quantum mechanical transition state energy and the index *j* loops over unique transition states leading to enantiomer *i*. Rearrangement of (1.3) gives an expression to compute ΔG^{TS} from the ee-value and *vice versa*.

$$\Delta G^{TS} = RT \ln\left(\frac{1+ee}{1-ee}\right) \tag{1.8}$$

For systems involving dioxirane catalysts and olefins with no rotational symmetry, eight geometrically distinct transition states were considered, as shown in Figure 2.4. For systems involving the C_{2h} symmetrical *trans*-stilbene (olefin **c**), the number of unique transition states is reduced by a factor of two. The same is true for systems involving the C_2 symmetrical catalyst **10**.

To locate transition states in their lowest energy conformations, initial guesses for all transition states were first obtained using the spiro UB3LYP/6-31G* epoxidation transition state of dimethyldioxirane and ethylene as a scaffold (Figure 2.3). While Shi invokes a planar transition state in some cases,[79-81] the planar transition state scaffold was not used, as Houk and coworkers have shown that the planar transition state for dimethyldioxirane and ethylene is 7.4 kcal/mol higher in energy than the corresponding spiro transition state at the B3LYP/6-31G* level, and represents only a secondary saddle point.[63] We attempted to address the conformational flexibility of the systems by

performing conformational searches of the transition state guesses using MacroModel[82] and the OPLS 2001 force field[83] while keeping the Cartesian coordinates of the reactive atoms of the transition state guesses frozen (atom numbers 1-5 in Figure 2.3). For some systems involving catalysts **14-19** with relatively flexible ring systems, conformational searching was also performed with the spiro carbon atoms of the dioxirane (C-1 in Figure 2.3) unfrozen, therefore allowing rotational adjustment of the transition state geometry, from strictly spiro to hybrid spiro-planar. This proved to be necessary as we observed many transition states that were neither completely planar, nor completely spiro, but rather a hybrid. Finally, where we suspected that the above did not yield transition states in their lowest energy conformations, we further subjected these structures to manual conformational searching.

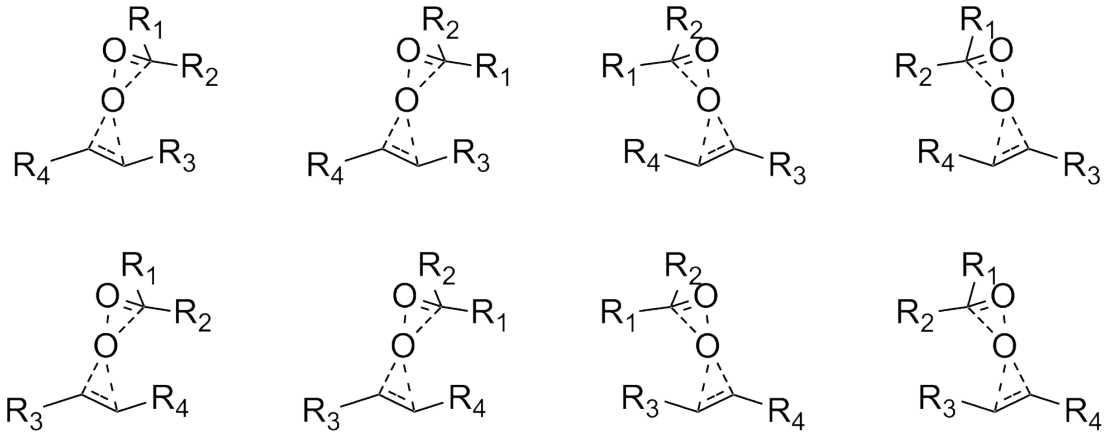


Figure 2.4. All unique dioxirane epoxidation transition states. For systems where either the catalyst or the olefin starting material contains rotational symmetry, fewer unique transition state geometries were considered.

We further refined our initial transition state guesses by performing minimum energy structure optimizations at the B3LYP/6-31G* level, while keeping three distances (C1-

O3, C5-O3 and C4-O3 in Figure 2.3a) between the reactive atoms fixed. The quantum mechanically refined transition state guesses were then used in all subsequent transition state optimizations.

It has already been shown that dioxirane-mediated epoxidations are well described (at least qualitatively) by B3LYP.[5,62-69] Houk *et al.*[63] have suggested that the transition states for epoxidation by dioxiranes lie in shallow potential energy wells, and that asymmetrical substitutions on the symmetrical dimethyldioxirane and ethylene transition state scaffold lead to more asynchronous transition states. Furthermore, they also showed that spin-unrestricted B3LYP (UB3LYP) leads to more asynchronous transition states than spin-restricted B3LYP. Therefore we employed unrestricted density functional theory within Jaguar 7.0 for all quantum mechanical computations to capture the asynchronicity and flexibility of the transition states. Transition states were first located in vacuum and then subsequently in the solution phase using the PBF solvation model as implemented in Jaguar 7.0. Solvation effects for the vacuum transition states were approximated from single-point calculations in the continuum solvent. Throughout the text we will use B3LYP/6-31G*(solvation)//B3LYP/6-31G*(vacuum) to indicate single point calculations within the solvation model on the transition states optimized in vacuum. For CH_3CN/H_2O and dioxane/H_2O mixtures, dielectric constants were taken as those reported in experimental measurements.[84,85] Because no experimental dielectric constants were reported for the other solvent mixtures, DMM/CH_3CN/H_2O and DMM/DME/H_2O, the dielectric constant was approximated as the mole fraction weighted average over all dielectric constants of the pure solvents. The dielectric constants ultimately employed were 55, 47, 20, 69, and 73 for 3:2 CH_3CN:H_2O (v:v), 3:1 CH_3CN:H_2O, 2:1

dioxane/H_2O, 3:3:6.4 DMM:CH_3CN:H_2O, and 26:1:2.8 H_2O:DMM:DME solvent systems, respectively. The solvent probe radius was set to 1.40 Å for all calculations as this probe radius was shown to reproduce solvation energies accurately for pure water.[86] The first shell correction factor term and the solute cavity energy term available in Jaguar 7.0 were included for all solvation calculations (keywords icavity=2 and isurf=1). Single point calculations at the B3LYP/6-311+G**(solution)//B3LYP/6-31G*(solution) level were also performed in order to check the basis set dependence of the results. Finally, all B3LYP transition states for a representative subset of the dataset were fully reoptimized at the M06-2X[27]/6-31G*(solution) level employing Jaguar Version 7.6 with the goal of obtaining a more accurate treatment of dispersion interactions, for which B3LYP is known to perform poorly.[17-20]

The nature of all vacuum B3LYP transition states was verified with frequency calculations, yielding only one large imaginary frequency. Harmonic zero-point energy corrections obtained from the frequency calculations of the vacuum B3LYP transition states were applied to all transition state energies. Because second derivatives within the solvation model could only be calculated via an impractically expensive numerical procedure with the versions of Jaguar used, vibrational frequencies of the transition states in solution were not computed. Harmonic thermochemical and entropic corrections other than the zero-point energy, for which low frequency vibrational modes contribute the most, were not included because the harmonic approximation is highly inaccurate for low frequency vibrational modes.

2.3 Results and Discussions

Figure 2.5 shows our correlation between the computed and the experimental enantioselectivities. We found that ΔG^{TS} is systematically over-estimated and therefore scaled all ΔG^{TS}'s by a scaling factor of 2/3 accordingly before using ΔG^{TS} to calculate the enantiomeric excesses. Inspection of Figure 2.5 shows that our method is able to capture which enantiomer a system will prefer, except for a few cases where low enantioselectivity is predicted. Furthermore, it is clear that our method consistently and accurately captures whether a system will have low (0-50%), medium (50-85%) or high (>85%) enantioselectivity for the majority of the systems studied. Specifically, most systems with high experimental enantioselectivity also have high computed enantioselectivity. The same is true for both medium and low enantioselectivities, suggesting that our method can be used to model in a quantitative sense the enantioselectivity of other dioxirane-catalysed epoxidations and perhaps even other enantioselective reactions as well.

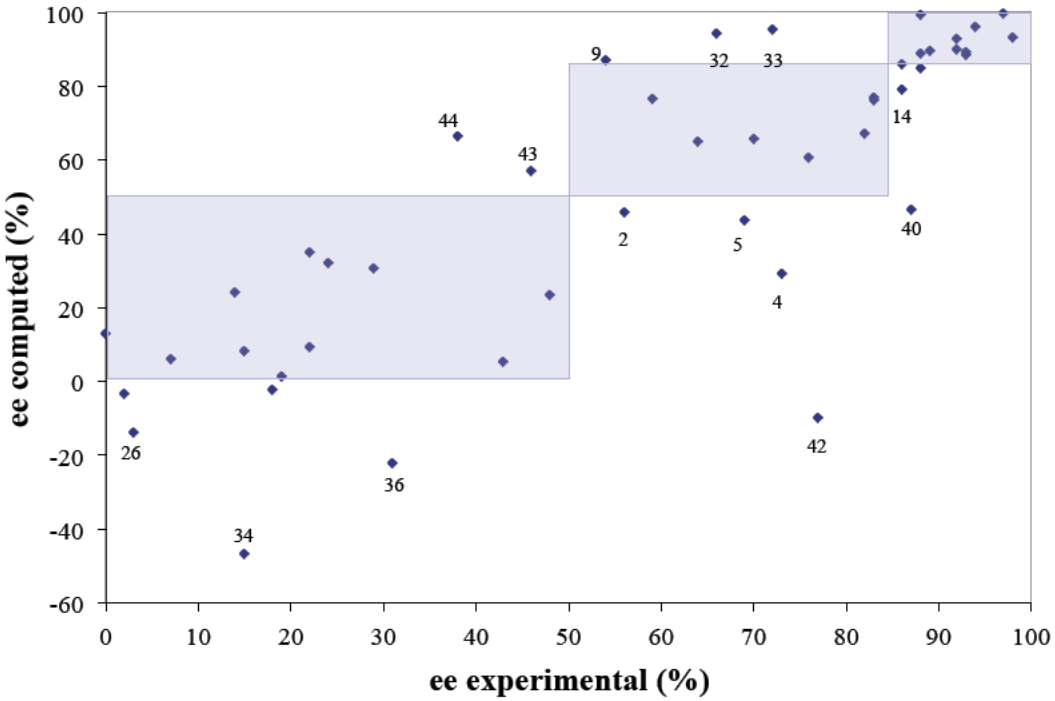


Figure 2.5. Experimental *vs.* computed % ee for all dioxirane catalyzed epoxidations. Results shown are for solution phase transition states at the B3LYP/6-31G*(solution) level. Shaded areas represent regions where computation and experiment fall within the same category of percent enantiomeric excess, i.e. low (0-50), medium (50-85) and high (85-100). Reactions that do not fall into the correct region are labeled with their corresponding reaction numbers. A linear least squares fit through all the points gives a coefficient of determination R^2 of 0.66.

The experimental and computed free energy differences, ΔG^{TS}, between the transition state ensembles leading to favored and disfavored enantiomers are plotted in Figure 2.6 for the B3LYP functional with the 6-31G* basis set. The results obtained with the M06-2X functional and the ones obtained with single point calculations with B3LYP in a larger basis set are shown in Figure 2.7 for comparison as well. (Detailed results at the B3LYP level with the 6-31G* basis are listed in Table 2.1 while the results obtained with the M06-2X functional and with single point B3LYP calculations in a larger basis set are shown in Table 2.2). Five outliers with deviations from experiment larger than 1.5 kcal/mol are present at the B3LYP/6-31G*(solution) level. As discussed in detail

later, all of these outliers share the unique structural motif of π-stacking between an oxygen or a chlorine atom and a benzene ring. Of course those outliers significantly distort the linear least squares fit shown in Figure 2.6b. Without the five outliers, the slope of the linear least squares fit becomes significantly closer to unity, with a value of 1.33 while the correlation coefficient, R^2, increases to a respectable value of 0.78.

Table 2.1. B3LYP/6-31G* results for dioxirane catalyzed epoxidations.

			ΔG^{TS} (kcal/mol) *(% ee)*							
Entry	Catalyst	Olefin	Exp.[1]		Vac.[2]		Soln.[3]		Ref.[4]	Solv.[5]
1	1	e	0.22	*(18)*	0.48	*(26)*	-0.05	*(-3)*	87	A
2	1	b	0.75	*(56)*	1.00	*(51)*	0.88	*(46)*	87	A
3	1	c	1.18	*(76)*	0.58	*(32)*	1.24	*(60)*	87	A
4	1	f	1.10	*(73)*	0.03	*(2)*	0.53	*(29)*	87	A
5	1	d	1.00	*(69)*	-0.07	*(-4)*	0.83	*(44)*	87	A
6	1	g	0.26	*(22)*	0.29	*(16)*	0.65	*(35)*	87	A
7	1	a	0.35	*(29)*	1.00	*(51)*	0.56	*(31)*	87	A
8	1	h	1.41	*(83)*	2.18	*(84)*	1.80	*(77)*	87	A
9	2	c	0.72	*(54)*	2.15	*(84)*	2.36	*(87)*	87	A
10	3	c	1.53	*(86)*	2.37	*(87)*	2.28	*(86)*	87	A
11	8	b	0.26	*(22)*	0.23	*(13)*	0.17	*(9)*	88	B
12	9	b	0.00	*(0)*	-0.40	*(-22)*	0.23	*(13)*	88	B
13	7	b	1.03	*(70)*	1.41	*(66)*	1.40	*(66)*	88	B
14	7	c	1.53	*(86)*	2.02	*(81)*	1.91	*(79)*	88	B
15	10	b	1.49	*(88)*	1.65	*(77)*	2.30	*(89)*	89	A
16	10	c	1.89	*(94)*	3.81	*(98)*	3.16	*(96.0)*	89	C
17	10	d	0.74	*(59)*	1.59	*(75)*	1.63	*(76)*	89	A
18	10	i	0.50	*(43)*	0.23	*(14)*	0.08	*(5)*	89	A
19	4	c	1.41	*(83)*	1.67	*(74)*	1.77	*(76)*	90	A
20	6	c	0.90	*(64)*	2.41	*(88)*	1.38	*(65)*	90	A
21	6	a	0.02	*(2)*	0.10	*(6)*	-0.06	*(-4)*	90	A
22	5	a	0.62	*(48)*	0.41	*(23)*	0.42	*(23)*	91	A
23	5	c	1.96	*(93)*	2.44	*(88)*	2.47	*(88)*	91	A
24	5	d	1.37	*(82)*	1.91	*(79)*	1.45	*(67)*	91	A
25	5	h	2.72	*(98)*	2.81	*(92)*	2.95	*(93)*	91	A
26	5	j	0.04	*(3)*	-0.60	*(-32)*	-0.25	*(-14)*	91	A
27	11	g	0.08	*(7)*	0.00	*(0)*	0.11	*(6)*	92	A
28	11	a	0.23	*(19)*	-0.22	*(-12)*	0.02	*(1)*	92	A
29	14	c	2.27	*(97)*	5.22	*(99.7)*	4.88	*(99.5)*	93	D
30	15	c	1.49	*(88)*	1.51	*(73)*	2.03	*(85)*	93	D
31	16	c	1.54	*(89)*	3.93	*(98.4)*	2.36	*(90)*	93	D
32	18	c	0.86	*(66)*	3.27	*(96.5)*	2.86	*(94)*	93	D
33	19	c	0.98	*(72)*	3.41	*(97.0)*	3.03	*(95)*	93	D
34	14	a	0.16	*(15)*	-1.00	*(-56)*	-0.79	*(-47)*	93	D
35	15	a	0.16	*(15)*	0.27	*(17)*	0.13	*(8)*	93	D
36	16	a	0.34	*(31)*	0.61	*(37)*	-0.35	*(-22)*	93	D
37	17	a	0.26	*(24)*	0.62	*(38)*	0.52	*(32)*	93	D
38	18	a	0.15	*(14)*	0.18	*(12)*	0.38	*(24)*	93	D
39	14	b	1.66	*(92)*	3.68	*(98.2)*	2.57	*(93)*	93	D
40	15	b	1.39	*(87)*	0.84	*(49)*	0.79	*(47)*	93	D
41	16	b	1.44	*(88)*	4.18	*(99.0)*	4.32	*(99.2)*	93	D
42	17	b	1.07	*(77)*	0.61	*(37)*	-0.16	*(-10)*	93	D
43	18	b	0.52	*(46)*	3.50	*(97.7)*	1.02	*(57)*	93	D
44	19	b	0.42	*(38)*	0.84	*(49)*	1.25	*(66)*	93	D
45	13	k	1.66	*(92)*	0.66	*(40)*	2.29	*(90)*	94	E
46	12	l	1.73	*(93)*	0.95	*(54)*	2.25	*(89)*	94	E

[1] Experimental ΔG^{TS}. [2] Results at the B3LYP/6-31G*(solution)//B3LYP/6-31G*(vacuum) level. The ee values were obtained from scaled ΔG^{TS}. (Scaling Factor = 2/3). [3] Results at the B3LYP/6-31G*(solution) level. The ee values were obtained from scaled ΔG^{TS}. (Scaling Factor = 2/3). [4] Experimental reference. [5] Solvent: (A) 3:2 CH_3CN:H_2O (v:v), dielectric constant, ε, employed in the continuum solvation model: 55 (B) 2:1 dioxane:H_2O (v:v), ε = 20 (C) 3:1 CH_3CN:H_2O (v:v), ε = 47 (D) 3:3:6.4 DMM:CH_3CN: H_2O (v:v:v), ε = 69 (E) 26:1:2.8 H_2O:DMM:DME (v:v:v), ε = 73.

Table 2.2. M06-2X/6-31G* and B3LYP/6-311+G** results for dioxirane epoxidations.

			ΔG^{TS} (kcal/mol) *(% ee)*							
Entry	Catalyst	Olefin	Exp.[1]		M06-2X[2]		larger basis[3]		Ref.[4]	Solv.[5]
1	1	e	0.22	*(18)*	-1.37	*(-65)*	-0.17	*(-9.4)*	87	A
2	1	b	0.75	*(56)*	0.97	*(50)*	0.23	*(13.0)*	87	A
3	1	c	1.18	*(76)*	1.02	*(52)*	1.01	*(51.5)*	87	A
4	1	f	1.10	*(73)*	-1.91	*(-79)*	-0.18	*(-10.1)*	87	A
5	1	d	1.00	*(69)*	0.70	*(37)*	1.15	*(57.0)*	87	A
6	1	g	0.26	*(22)*	0.60	*(32)*	0.15	*(8.6)*	87	A
7	1	a	0.35	*(29)*	0.82	*(43)*	0.41	*(22.7)*	87	A
8	1	h	1.41	*(83)*	3.28	*(95.1)*	0.94	*(48.4)*	87	A
9	2	c	0.72	*(54)*	0.93	*(48)*	2.44	*(87.9)*	87	A
10	3	c	1.53	*(86)*	-0.96	*(-49)*	1.38	*(65)*	87	A
11	8	b	0.26	*(22)*	-0.09	*(-5)*	0.61	*(33)*	88	B
12	9	b	0.00	*(0)*	0.37	*(21)*	-0.24	*(-13)*	88	B
13	7	b	1.03	*(70)*	1.95	*(80)*	0.93	*(48)*	88	B
14	7	c	1.53	*(86)*	2.26	*(86)*	1.77	*(76)*	88	B
15	10	b	1.49	*(88)*	5.01	*(99.6)*	1.20	*(63)*	89	A
16	10	c	1.89	*(94)*	6.14	*(99.9)*	3.08	*(96)*	89	C
17	10	d	0.74	*(59)*	5.01	*(99.6)*	1.10	*(59)*	89	A
18	10	i	0.50	*(43)*	3.85	*(98.3)*	-0.04	*(-3)*	89	A
19	4	c	1.41	*(83)*	0.82	*(43)*	1.69	*(74)*	90	A
20	6	c	0.90	*(64)*	8.68	*(99.9)*	1.26	*(61)*	90	A
21	6	a	0.02	*(2)*	0.73	*(39)*	0.26	*(15)*	90	A
22	5	a	0.62	*(48)*	0.67	*(36)*	0.39	*(22)*	91	A
23	5	c	1.96	*(93)*	-1.30	*(-63)*	1.69	*(74)*	91	A
24	5	d	1.37	*(82)*	-1.47	*(-68)*	0.86	*(45)*	91	A
25	5	h	2.72	*(98)*	2.16	*(84)*	1.86	*(78)*	91	A
26	5	j	0.04	*(3)*	-0.68	*(-36)*	0.78	*(41)*	91	A
27	11	g	0.08	*(7)*	na	*(na)*	-0.59	*(-32)*	92	A
28	11	a	0.23	*(19)*	na	*(na)*	-0.55	*(-30)*	92	A
29	14	c	2.27	*(97)*	6.49	*(99.9)*	4.98	*(99.6)*	93	D
30	15	c	1.49	*(88)*	4.41	*(99.1)*	2.23	*(88)*	93	D
31	16	c	1.54	*(89)*	na	*(na)*	1.65	*(77)*	93	D
32	18	c	0.86	*(66)*	1.54	*(74)*	2.38	*(90)*	93	D
33	19	c	0.98	*(72)*	2.23	*(87.9)*	2.78	*(93.6)*	93	D
34	14	a	0.16	*(15)*	0.87	*(50)*	-1.59	*(-77)*	93	D
35	15	a	0.16	*(15)*	-0.63	*(-38)*	0.21	*(13)*	93	D
36	16	a	0.34	*(31)*	na	*(na)*	0.41	*(26)*	93	D
37	17	a	0.26	*(24)*	na	*(na)*	0.29	*(18)*	93	D
38	18	a	0.15	*(14)*	0.87	*(50)*	0.13	*(8)*	93	D
39	14	b	1.66	*(92)*	2.63	*(93.3)*	0.74	*(44)*	93	D
40	15	b	1.39	*(87)*	3.27	*(97.0)*	-0.15	*(-9)*	93	D
41	16	b	1.44	*(88)*	na	*(na)*	4.57	*(99.4)*	93	D
42	17	b	1.07	*(77)*	6.30	*(99.9)*	-1.99	*(-85)*	93	D
43	18	b	0.52	*(46)*	1.93	*(84)*	0.43	*(27)*	93	D
44	19	b	0.42	*(38)*	1.40	*(71)*	0.68	*(41)*	93	D
45	13	k	1.66	*(92)*	na	*(na)*	2.84	*(94.8)*	94	E
46	12	l	1.73	*(93)*	na	*(na)*	0.45	*(28)*	94	E

[1] Experimental ΔG^{TS}. [2] Results at the M06-2X/6-31G*(solution) level. Zero point energies were obtained from the B3LYP/6-31G*(vacuum) frequency calculations. The ee values were obtained from scaled ΔG^{TS}. (Scaling Factor = 2/3). [3] Results at the B3LYP/6-311+G**(solution)//B3LYP/6-31G*(solution) level. The ee values were obtained from scaled ΔG^{TS}. (Scaling Factor = 2/3). [4] Experimental reference. [5] Solvent: (A) 3:2 CH_3CN:H_2O (v:v), dielectric constant, ε, employed in the continuum solvation model: 55 (B) 2:1 dioxane:H_2O (v:v), ε = 20 (C) 3:1 CH_3CN:H_2O (v:v), ε = 47 (D) 3:3:6.4 DMM:CH_3CN: H_2O (v:v:v), ε = 69 (E) 26:1:2.8 H_2O:DMM:DME (v:v:v), ε = 73.

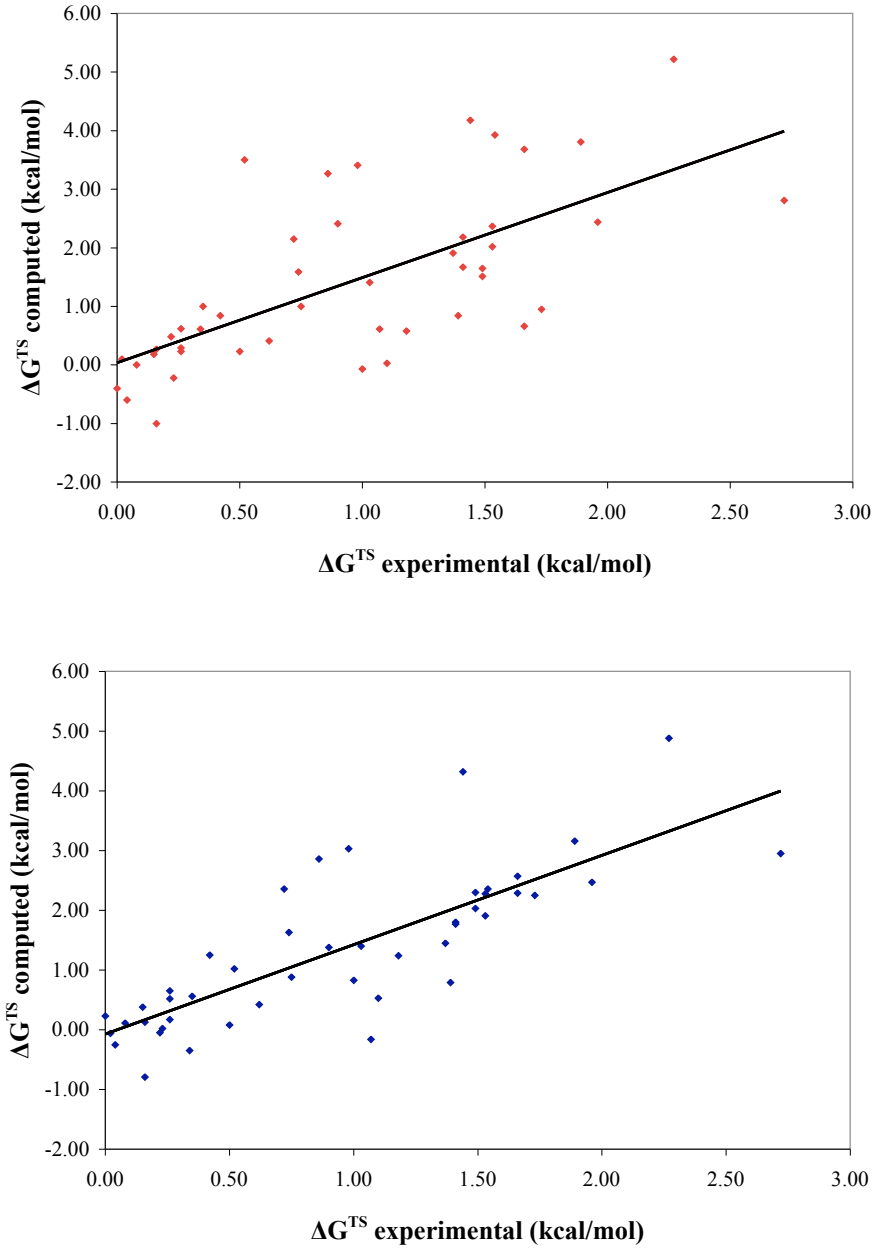


Figure 2.6. Experimental *vs.* computed (B3LYP/6-31G*) transition state free energy differences (ΔG^{TS}) for all dioxirane catalyzed epoxidations. **(a)** B3LYP/6-31G*(solution)// B3LYP/6-31G*(vacuum) level. (The solid line shows a linear least squares fit of y = 1.45x + 0.04 with coefficient of determination $R^2 = 0.46$). (b) B3LYP/6-31G*(solution) level (y = 1.49x - 0.07, R2 = 0.64).

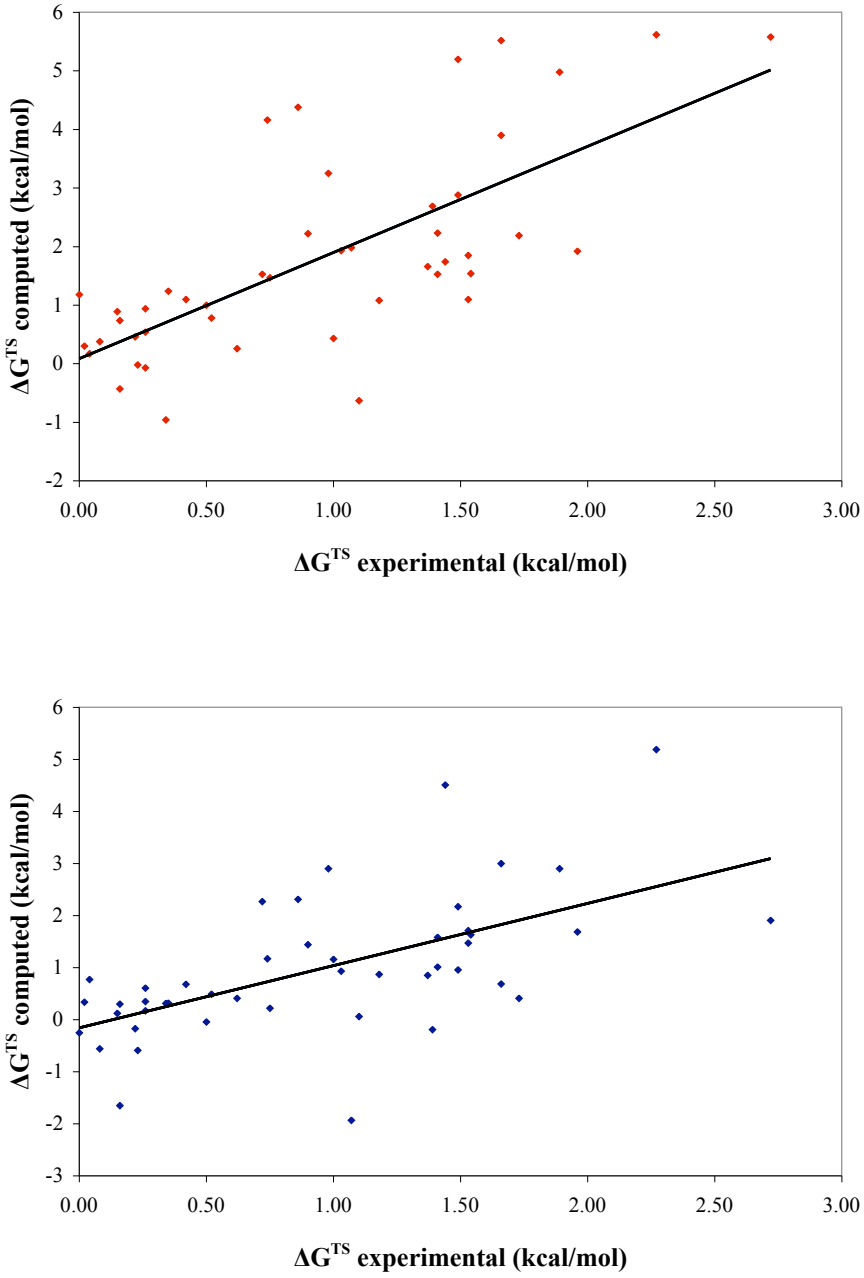
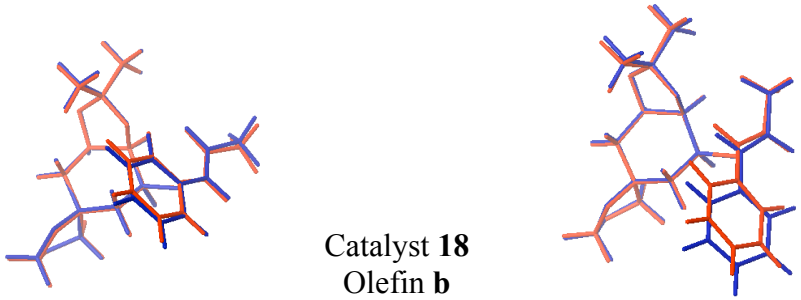


Figure 2.7. Experimental *vs.* computed (M06-2X/6-31G* and B3LYP/6-311+G**) transition state free energy differences (ΔG^{TS}) for all dioxirane catalyzed epoxidations. (a) M06-2X/6-31G*(solution) level. (The solid line shows a linear least squares fit of y = 1.29x + 0.61 with coefficient of determination R^2 = 0.12). (b) B3LYP/6-311+G**(solution)//B3LYP/6-31G*(solution) level (y = 1.19x - 0.16, R^2 = 0.36).

Interestingly, a comparison of Figure 2.6a and Figure 2.6b shows a significant improvement in correlation with the experimental results when the transition states are optimized in the solution phase, rather than simply estimating solvation energies with single point calculations, with correlation coefficients of 0.46 and 0.64, respectively. This improvement is unsurprising given that all systems treated were aqueous, suggesting that their transition state geometries might change significantly by reoptimizing the vacuum transition states within the solvation model. This strategy differs from previous work in which gas phase optimized transition states were employed exclusively under the assumption that the transition state geometries would not vary drastically from vacuum to solution phase.[95-97] Our results show the dubious nature of this common assumption. Figure 2.8 shows the lowest energy favored and disfavored transition states in vacuum and solution phase for cases, which show the most improvement following reoptimization in solution phase. Notably, the two forming bonds (O3-C4 and O3-C5 in Figure 2.3) are appreciably longer when the transition states are located in solution rather than vacuum phase. In order to quantify this effect we defined average forming bond lengths as a the Boltzmann weighted average of the corresponding bond lengths of all transition structures. The computed average bond lengths for all reactions are listed in Table 2.3. Each asynchronous transition state is characterized by two forming average bond lengths, one being shorter then the other. In solution phase these bond lengths are 2.40 ± 0.02 Å (mean ± standard error (s.e.); sample size $n = 46$) and 2.15 ± 0.01 Å (mean ± s.e.; $n = 46$) respectively. This is longer than the corresponding averaged bonds lengths of the vacuum transition states, which are 2.30 ± 0.01 Å (mean ± s.e.; $n = 46$) and 1.97 ± 0.01 Å (mean ± s.e.; $n = 46$) respectively.

Lowest energy *favored* transition state Lowest energy *disfavored* transition state

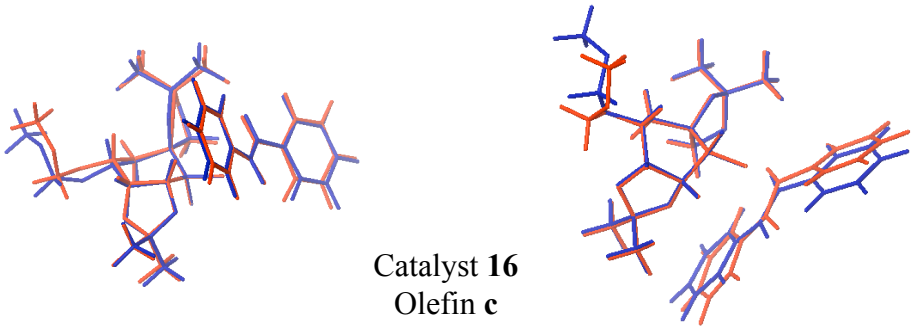
(a)



Catalyst **18**
Olefin **b**

43g Improvement 2.5 kcal/mol 43a

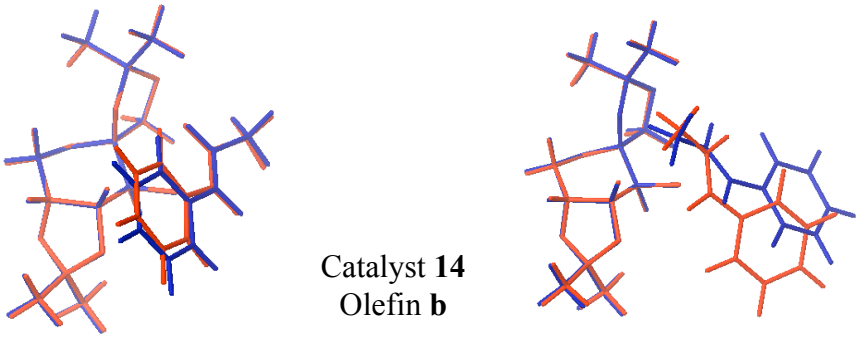
(b)



Catalyst **16**
Olefin **c**

31a Improvement 1.6 kcal/mol 31d

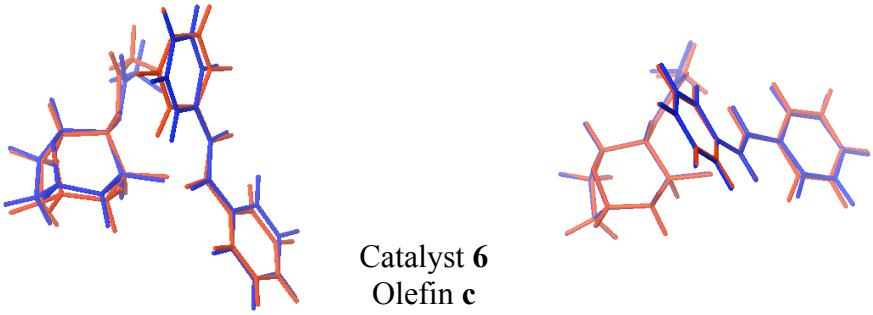
(c)



Catalyst **14**
Olefin **b**

39a Improvement 1.1 kcal/mol 39f

(d)



Catalyst **6**
Olefin **c**

20b Improvement 1.0 kcal/mol 20d

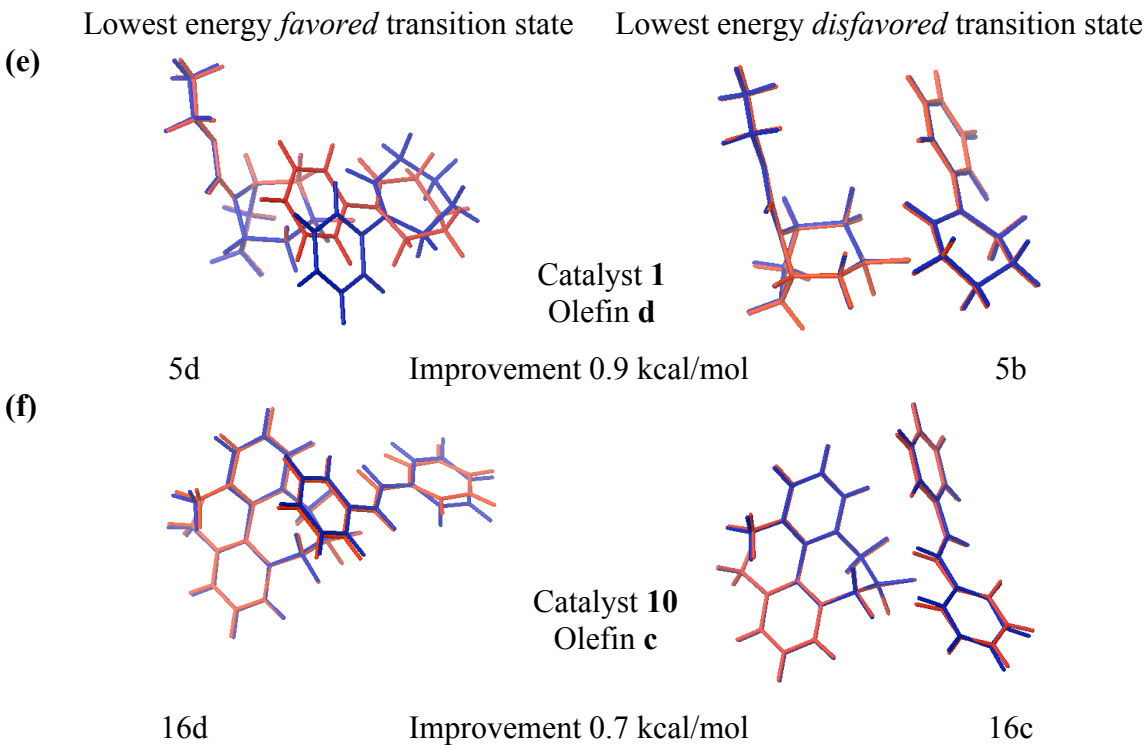


Figure 2.8. Structural changes of the dioxirane epoxidation transition states due to optimization in solvent. The lowest energy B3LYP/6-31G*(vacuum) transition state structures (red) are superimposed on the B3LYP/6-31G*(solution) transition state structures (blue) for reactions 43 **(a)**, 31 **(b)**, 39 **(c)**, 20 **(d)**, 5 **(e)**, 16 **(f)**, which show the most significant improvements following reoptimization of the transition states in solution phase. Reductions in absolute deviations from experimental ΔG^{TS} are also shown.

Table 2.3. Asymmetry in the dioxirane epoxidation transition states. Boltzmann weighted forming bond distances (O3-C4 and O3-C5 in Figure 2.3) in Å of the transition states at the B3LYP/6-31G*(solution)//B3LYP/6-31G*(vacuum) and the B3LYP/6-31G*(solution) levels are shown.

	vacuum structures		solution phase structures	
Reaction #	**shorter**	**longer**	**shorter**	**longer**
1	2.03	2.37	2.22	2.48
2	1.97	2.33	2.16	2.44
3	1.99	2.24	2.20	2.31
4	2.00	2.29	2.19	2.38
5	2.02	2.38	2.21	2.48
6	1.98	2.46	2.17	2.59
7	1.96	2.39	2.14	2.50
8	2.00	2.39	2.11	2.42
9	2.01	2.11	2.21	2.22
10	1.95	2.20	2.12	2.29
11	1.90	2.27	2.06	2.35
12	1.93	2.25	2.12	2.37
13	1.96	2.33	2.14	2.43
14	1.96	2.23	2.14	2.30
15	1.98	2.30	2.24	2.41
16	2.01	2.12	2.25	2.29
17	2.02	2.34	2.24	2.53
18	1.93	2.37	2.13	2.47
19	1.99	2.19	2.20	2.31
20	2.03	2.26	2.18	2.42
21	2.02	2.42	2.19	2.56
22	1.90	2.36	2.08	2.48
23	2.00	2.08	2.16	2.22
24	1.97	2.33	2.14	2.47
25	1.97	2.46	2.12	2.55
26	1.89	2.15	2.05	2.11
27	1.94	2.42	2.17	2.58
28	1.94	2.36	2.14	2.52
29	2.00	2.11	2.11	2.13
30	1.94	2.21	2.00	2.23
31	1.99	2.12	2.12	2.21
32	2.01	2.21	2.23	2.34
33	2.02	2.21	2.22	2.25
34	1.94	2.38	2.14	2.52
35	1.93	2.39	2.11	2.51
36	1.94	2.40	2.12	2.53
37	1.93	2.37	2.14	2.51
38	1.97	2.36	2.16	2.48
39	1.99	2.27	2.22	2.44
40	1.93	2.29	2.13	2.37
41	1.95	2.38	2.14	2.46
42	1.94	2.33	2.05	2.30
43	1.91	2.30	2.21	2.43
44	2.00	2.35	2.19	2.42
45	1.94	2.38	1.92	2.37
46	1.97	2.22	2.19	2.53

While locating transition states in the solution phase leads to significantly improved results, five outliers persist. Specifically, cases 41, 29, 33, 32 and 9 (listed in decreasing order of error) are not treated well by our model in either vacuum or solution phase and show deviations from experimental ΔG^{TS} larger than 1.5 kcal/mol at the B3LYP/6-31G*(solution) level. All outliers share one common characteristic: the relevant disfavored transition states have either an oxygen or a chlorine atom stacked over the face of a benzene ring (as illustrated in Figure 2.9a). This structural motif seems to be present only for cases with large discrepancies between computation and experiment, but is absent for all other cases. For each ensemble of favored and disfavored B3LYP/6-31G* solution phase transition states, the Boltzmann weighted average of the number of oxygen, fluorine or chlorine atoms present in a cone above the plain of each benzene ring was computed. The differences between these average values of disfavored and favored transition state ensembles were then plotted against the deviation of the computed ΔG^{TS} from experimental values (Figure 2.9b).

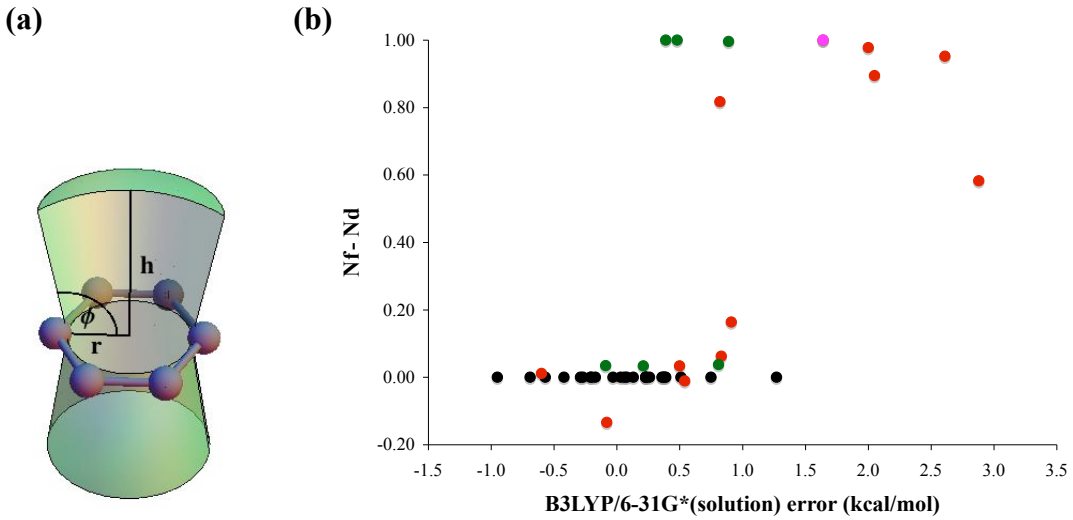
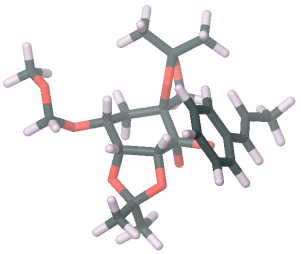
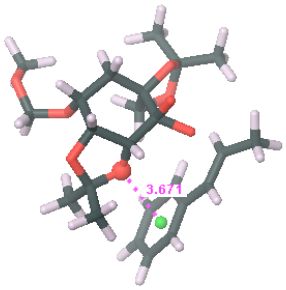
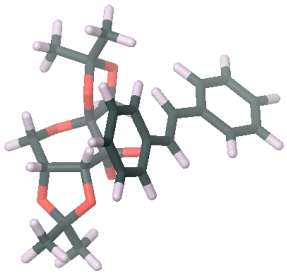
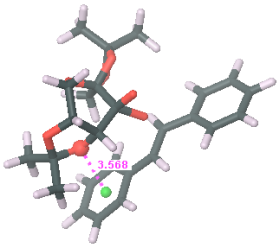
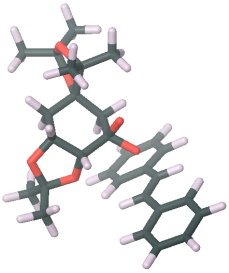
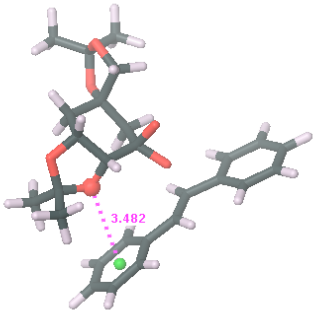
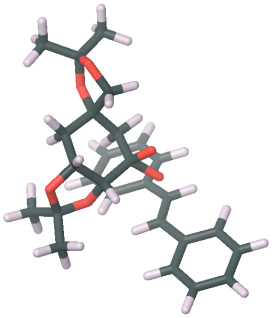
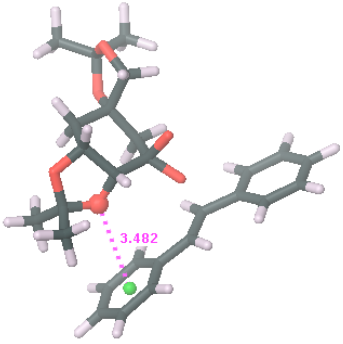


Figure 2.9. Analysis of potential reasons for the outliers in the dioxirane epoxidation dataset. **(a)** Schematic illustration of the cone volume above benzene rings searched for oxygen, fluorine and chlorine atoms. Benzene carbon atoms are shown in gray. h = 4.0 Å, r = 1.5 Å and ϕ = 100°. **(b)** Difference between disfavored and favored transition state ensembles in the total number of oxygen, fluorine and chlorine atoms (Boltzmann weighted average) present within the cone shown in Figure 2.9a above each benzene ring plotted against deviation from experiment. black = no atom in cone, red = oxygen in cone, green = fluorine in cone and pink = chlorine in cone.

Based on the well known deficiencies of B3LYP in treating dispersion and especially π-stacking interactions,[17-20] it is not surprising to see that π-stacking interactions with oxygen and chlorine pose an obstacle for our computational model. Interestingly, π-stacking interactions with the smaller and less polarizable fluorine seem to be treated much more accurately by our model than such interactions involving oxygen or other halogens. Because B3LYP generally does not take account of attractive π-stacking interactions, an inflated computed transition state free energy gap could be the result of an artificial energy increase of the disfavored transition state ensemble due to repulsive B3LYP errors for π-stacking interactions. Where π-stacking interactions with oxygen or chlorine are absent from a reaction, dispersion related B3LYP errors are

presumably minimized due to fortuitous error cancellation. However, since B3LYP errors for π-stacking interactions are often several kcal/mol higher than the corresponding errors for other types of dispersion interactions, differential error likely accumulates where strong π-stacking interactions are present. This is illustrated by comparing the B3LYP/6-31G* errors for the stacked (C2h) and the T-shaped (C2v) benzene dimers. Single point B3LYP/6-31G* calculations on *ab initio* geometries reported by Urecka and coworkers[98] give binding energies of 2.58 and 0.18 kcal/mol for the stacked and the T-shaped benzene-dimers, respectively, while best estimates[98] for these interaction energies are -2.73 and -2.74 kcal/mol. Therefore B3LYP seems to give the largest repulsive errors for systems showing strong π-stacking interactions.

	Lowest energy favored transition state		Lowest energy disfavored transition state
(a)		Catalyst **16** Olefin **b**	
	41h (0.0)	Error: 2.9 kcal/mol	41c (4.5)
(b)		Catalyst **14** Olefin **c**	
	29a (0.0)	Error: 2.6 kcal/mol	29d (4.9)
(c)		Catalyst **19** Olefin **c**	
	33c (0.0)	Error: 2.1 kcal/mol	33d (3.1)
(d)		Catalyst **18** Olefin **c**	

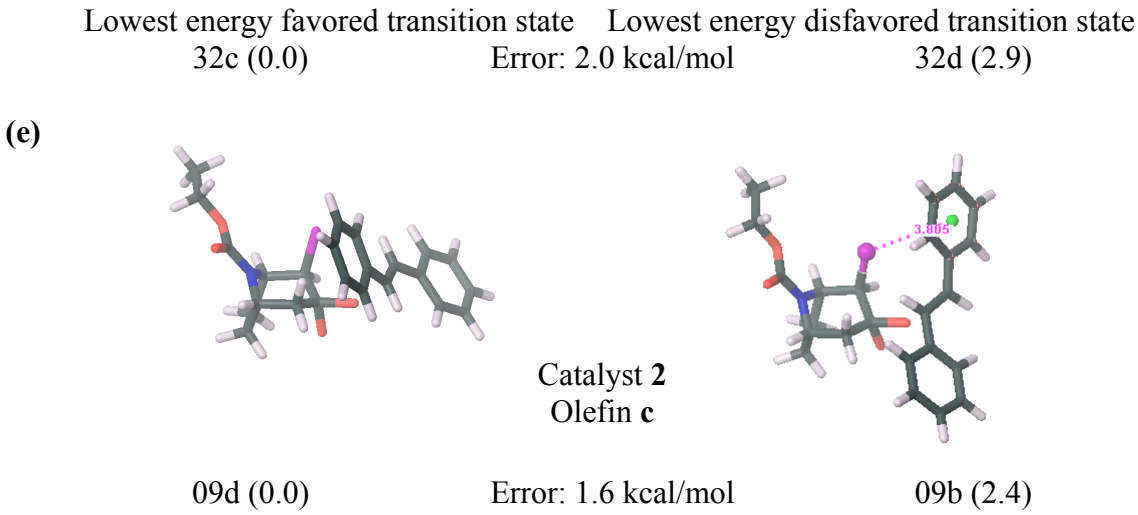


Figure 2.10. Structures of dioxirane epoxidation transition states with large errors. The lowest energy favored and disfavored transition states at the B3LYP/6-31G* (solution) level are shown for all cases where the error at this level is larger than 1.5 kcal/mol. Deviations (in kcal/mol) from the experimental transition state free energy differences, ΔG^{TS}, are given for each reaction. Relative energies of the transition states (in kcal/mol) are shown in brackets. π-stacking interaction distances between benzene middle points (green) and oxygen or chlorine atoms (in Å) are shown. Atom colors: carbon (black), oxygen (red), hydrogen (white), nitrogen (blue) and chlorine (pink). **(a)** reaction 41, **(b)** reaction 29, **(c)** reaction 33, **(d)** reaction 32, **(e)** reaction 9.

In order to confirm that the incorrect treatment of dispersion interactions by B3LYP is indeed a major source of error in our dataset, all transition states of a representative subset were reoptimized at the M06-2X[27]/6-31G*(solution) level. Truhlar's M06-2X functional performs well for dispersion dominated complexes as well as for activation energies.[28] However, the preliminary results obtained with M06-2X on our dataset are inferior to the B3LYP results. (For a comparison of the statistical performances of all methods employed in this study see Table 2.2.) The mean unsigned error as well as the correlation between experimental and computed ΔG^{TS} is clearly worse for M06-2X than for B3LYP. It is noteworthy, however, that three of the five B3LYP/6-31G* (solution) outliers improve significantly, namely systems 9, 32 and 33. For those systems the errors with respect to experiment are reduced by 1.4, 1.3 and 0.8

kcal/mol respectively with the M06-2X optimizations. However, the B3LYP outlier 29 gets significantly worse (error increases by 1.6 kcal/mol) when optimized with M06-2X while for the fifth outlier, 41, M06-2X results are not available due to convergence problems for some of the transition states.

The M06-2X results are somewhat surprising, in view of the strong performance demonstrated by this functional to data on a variety of training sets, and for a wide range of properties. There are several possible explanations for the results, which we plan to pursue in future work:

(1) Our solvation model has not been reoptimized for M06-2X; we simply used the model previously optimized for B3LYP. Hence if the charge distributions obtained with M06-2X are sufficiently different (and that might not be an enormous quantitative shift, in view of the small energy differences we are talking about), this could lead to degradation of the relative energetics.

(2) We spent significantly less time creating the optimized transition states for M06-2X than we did when working on the data set with B3LYP. There may for example be alternative minima with the former functional that we did not explore. Again, this could lead problems for specific compounds where one or both of the transition state structures are not in the lowest free energy state.

(3) While M06-2X has displayed encouraging performance for a number of training sets, these data sets are relatively small, and do not cover a broad range of chemical problems. M06-2X also contains a large number of fitting parameters (on the order of 35) as compared to the 3 parameters in B3LYP. Hence, there has to be some concern about overfitting (which would yield superior results for the training set as compared to what

one might expect in an application not covered by those data sets) and what we have observed may be one example of this type.

Table 2.4. Performance of various methods for the dioxirane epoxidation dataset.

Method	MUE[1]	Max Error[2]	R^3
B3LYP/6-31G*(solv.)//B3LYP/6-31G*(vac.)	0.85	3.0	0.46
B3LYP/6-31G*(solv.)	0.65	2.9	0.64
B3LYP/6-311+G**(solv.)//B3LYP/6-31G*(solv.)	0.73	3.1	0.36
M06-2X/6-31G*(solv.)[4]	1.76	5.2	0.12

[1] Mean unsigned error (kcal/mol) from experimental ΔG^{TS}.
[2] Maximum absolute deviation (kcal/mol) from experimental ΔG^{TS}.
[3] Correlation coefficient for a linear least squares fit to a plot of computed versus experimental ΔG^{TS}.
[4] Due to problems in converging some of the transition state geometries, systems 27, 28, 31, 36, 37, 41, 45, 46 are not included in the M06-2X dataset.

In addition to the M06-2X calculations, single point calculations of the solution phase transition states at the B3LYP/6-311+G**(solution)//B3LYP/6-31G*(solution) level were also performed to probe basis set dependence. Table 2.2 shows that the correlation coefficient becomes notably worse compared to the results obtained with the smaller basis set. This could simply be due to random errors arising from the fact the geometries are not actually optimized within the larger basis set. Singleton has shown that relative energies at the B3LYP/6-311+G**(vacuum)//B3LYP/6-31G*(vacuum) level vary only by about 0.2 kcal/mol from the energies at the B3LYP/6-311+G**(vacuum)//B3LYP/6-31+G**(vacuum) level for two Shi epoxidations.[68] However changes in geometry due to changes in basis set might be larger for transition states optimized in solution phase since the solvation energies also depend on the basis sets used. In addition, since we seek to discriminate free energy differences within a range of only about 2.5 kcal/mol, even small errors in relative energies, which are likely to be negligible for other problems, lead to a dramatic degradation of the correlation with experiment.

Notably, four out of five outliers of the dataset are epoxidations with Shi's catalyst[33] and it is therefore not surprising that the MUE for the Shi epoxidations (entries 29 – 46) is significantly higher than for the epoxidation reactions using other dioxirane catalysts (entries 1 – 28), with values of 1.02 and 0.41 kcal/mol at the B3LYP/6-31G*(solvent) level, respectively. Inaccurate treatment of dispersion interactions by B3LYP is one likely reason for this discrepancy. However, additional factors could further explain the computational model's difficulty in modeling the Shi epoxidation in particular.

First, dioxirane-catalyzed epoxidations have been shown to be particularly sensitive to solvent selection, presumably because the transition state exhibits enhanced hydrogen bonding.[63] In cases where the olefin contains hydrophobic groups (such as phenyl substituents), hydrophobic effects likely also influence the enantioselectivity. However, continuum solvation models might not accurately capture such solvation effects. Notably, the sugar-derived dioxirane catalysts employed by Shi (catalysts **12 – 19**) contain many potential hydrogen bond acceptors and the corresponding reactions are also performed in highly polar solvent systems. Therefore, errors due to approximating the solvent with a dielectric continuum are likely to be largest for the Shi epoxidation reactions. A plot of MUE against the number of hydrogen bond acceptors can be found in Figure 2.11.

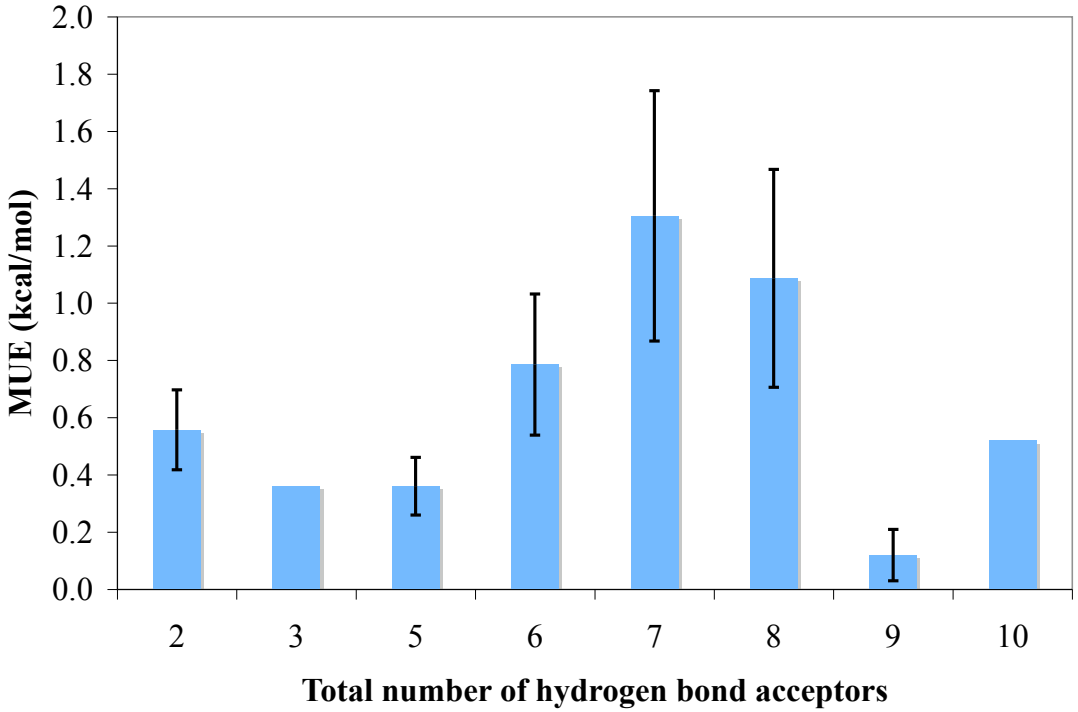


Figure 2.11. Effects of hydrogen bonding on the prediction accuracy for dioxirane catalyzed epoxidations. The plot shows mean unsigned error (MUE) as a function of the total number of hydrogen bond acceptors present in the transition states. Only Oxygen and Nitrogen were counted as hydrogen bond acceptors. n = sample size. Error bars represent standard error. No error bars are shown for n = 1.

Secondly, as noted by Singleton and Houk,[5,62-65,67,68,97] the Shi epoxidation can have many conformationally distinct low-energy transition states. To quantify the degree of conformational flexibility for each reaction, each catalyst and olefin was subjected to conformational searching within MacroModel using the OPLS 2001 force field employing 20,000 steps. The sum of unique conformations of catalyst and olefin conformations within 5 kcal/mol was then taken as a measure of the conformational flexibility of a system. Accordingly we found that the Shi epoxidation catalysts (entries 29 – 46 in Table 2.1) show a distinctly higher degree of conformational flexibility

compared to the rest of the dataset with values of 52 ± 11 (mean ± s.e.; n = 18) and 17 ± 11 (mean ± s.e.; n = 28) respectively. Even though we tried to search conformational space as completely as possible according to our protocol described in the methods section, we cannot exclude the possibility that for some flexible systems there might exist low energy conformations of transition states not considered in this study. Therefore, part of the observed errors for the more flexible Shi epoxidations could also results from incomplete conformational searching.

Much debate within the literature has centered around whether the dioxirane-mediated epoxidation proceeds through a spiro (Figure 2.3a) or planar (Figure 2.3b) transition state. Computational studies at various levels of theory have shown the preference of spiro over planar transition states for small model systems.[99] However, Shi has proposed that some reactions with sterically hindered transition states as well as reactions involving *cis*-olefins proceed through planar transition states.[79-81] In agreement with the results obtained by Singleton,[68] the results of our work indicate that all relevant low energy transition states are either spiro or hybrid spiro-planar. No relevant transition states with a completely planar geometry were found. In order to classify the degree of rotation in the transition states for each reaction, Boltzmann-weighted averages of the dihedral angles, θ, defined in Figure 2.12, were computed.

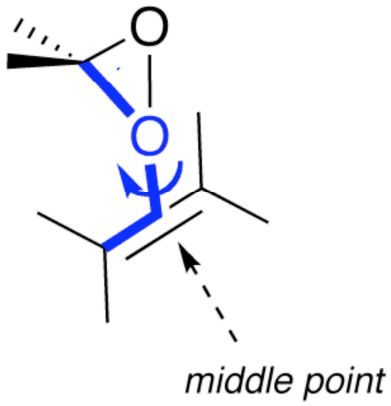


Figure 2.12. Definition of the dihedral angle θ (blue), which describes a smooth transition of dioxirane epoxidation transition states from "spiro" to "planar".

For each ensemble of favored and disfavored transition states the averaged dihedral angles θ of the solution phase transition states are listed in Table 2.5. Interestingly, dihedral angles for the *disfavored* transition states deviate more from 90° (which corresponds to a perfect spiro transition state) than do the dihedral angles for the *favored* transition states, with average values for the whole dataset of 117 ± 2° (mean ± s.e.; n = 46) and 100 ± 1° (mean ± s.e.; n = 46) respectively. This suggests that the disfavored transition states are often distorted from spiro in order to reduce steric clashes. However, while Shi proposed that this distortion leads to planar transition states, we show that less distorted hybrid spiro-planar transition states result instead. In addition, the degree of rotation in the transition states seems to be at least partly responsible for the increased energy of the disfavored transition states as compared to the favored ones.

Table 2.5. "Spiro" *vs.* "planar" dioxirane epoxidation transition states. Boltzmann-weighted dihedral angles θ (defined in Figure 2.12) are shown at the B3LYP/6-31G*(solution) level for each ensemble of transition states.

	θ (Degrees)	
Reaction #	**Favored**	**Disfavored**
1	91.0	116.3
2	96.8	133.6
3	91.9	123.3
4	97.1	128.6
5	110.5	128.0
6	123.4	110.4
7	95.0	98.1
8	102.0	123.7
9	99.0	120.7
10	91.4	136.2
11	99.3	98.3
12	98.9	95.4
13	98.9	128.4
14	98.1	135.6
15	94.3	132.9
16	93.3	100.3
17	97.2	133.4
18	96.8	103.9
19	92.2	120.6
20	105.1	118.7
21	106.1	95.2
22	105.9	112.3
23	96.4	141.8
24	90.1	140.4
25	92.1	141.5
26	111.9	103.3
27	123.3	136.2
28	92.9	108.7
29	94.2	114.6
30	95.1	92.2
31	102.6	118.2
32	90.6	92.3
33	92.0	95.5
34	98.7	117.8
35	100.6	126.0
36	126.9	116.6
37	101.9	115.4
38	92.9	103.3
39	98.3	133.2
40	90.3	106.8
41	95.1	106.1
42	92.5	115.0
43	94.2	133.3
44	93.8	133.0
45	132.4	97.6
46	101.6	97.3

Other groups have also illustrated the importance of asynchronous transition structures.[68] Unsurprisingly, we find that the transition state asynchronicity is generally larger for asymmetrically substituted than for symmetrically substituted olefins. In order to quantify these effects, we defined asynchronicity as the absolute difference between the two forming bond lengths (O3-C4 and O3-C5 distances in Figure 2.3). Representative bond lengths for each reaction were then computed as the Boltzmann-weighted averages of the corresponding shorter and longer distances of all transition states. Asynchronicities for each reaction were computed as absolute differences between the Boltzmann-weighted distance averages. Table 2.6 shows mean asynchronicities for all olefins present in the dataset (ordered according to increasing asynchronicity). The electron deficient 2-cyclohexenone (olefin **j**) forms transition states with lowest asynchronicity. The symmetrically substituted *trans*-stilbene (olefin **c**) forms transition states with clearly lower asynchronicity than the other asymmetrically substituted olefins. Consistent with previous findings[68,100] a dioxirane-catalyzed epoxidation is an electrophilic process, i.e. we find a Boltzmann-weighted partial charge (Computed as the sum over all mulliken charges of the olefin atoms at the B3LYP/6-31G*(solution) level) on the olefin-part of the transition states with an average value of 0.38 ± 0.01 (mean $\pm$ s.e.; $n = 46$). Considering the electrophilic nature of the epoxidations, it is therefore not surprising that we find the longer of the two forming bonds always to be the one between the dioxirane-oxygen and the carbon α to conjugated substituents such as phenyl or alkene. The only exceptions to this rule are a few high-energy transition states, i.e. 41a, 41e, 42b (solution phase structures) as well as 39f, 40a, 41a, 41e, 42b (vacuum structures). Olefin h displays the longer forming bond as expected always between the dioxirane oxygen and the

carbon α to the *two* phenyl substituents. Olefin f, containing two symmetrically arranged phenyl substituents, as well as olefin j with very low asynchronicity both show no clear trends in which of the two forming bond lengths is shorter in the transition states.

Table 2.6. Asynchronicities for all olefins of the dioxirane epoxidation dataset.

Olefin	Asynchronicity (Å) (mean ± standard error)	sample size
j	0.07	1
c	0.11 ± 0.02	13
f	0.19	1
b	0.25 ± 0.01	11
e	0.26	1
d	0.30 ± 0.02	3
i	0.33	1
l	0.34	1
h	0.37 ± 0.06	2
a	0.38 ± 0.01	9
g	0.41 ± 0.01	2
k	0.45	1

In summary, the work described in this chapter represents the first fully quantum mechanical study of quantitative modeling of enantioselectivity for a large dataset. Our initial results obtained with dioxirane catalyzed epoxidations in aqueous solvents indicate that optimization of transition states within a solvation model is critical to the accurate and robust prediction of experimental results. With optimization in solvent, our newly developed methodology predicts the relative transition state free energy differences of enantiomeric transition states quite accurately (with a MUE < 0.5 kcal/mol) in combination with the B3LYP functional and the 6-31G* basis set. However, by carefully inspecting the outliers of our dataset, we also found that the well-known deficiencies of B3LYP in treating dispersion interactions are likely causing a significant reduction of the predictive accuracy of our methodology. Therefore, I and Dr. Arteum Bochevarov from

the Friesner lab developed corrections, which greatly improve the accuracy of B3LYP for non-covalent interactions and also partially correct for Basis Set Superposition Error (BSSE). These corrections will be described in the following chapter of this dissertation.

3 Improving Density Functional Theory for non-covalent interactions

3.1 Introduction

Density Functional Theory (DFT)[101] has become an indispensable computational method to solve real world problems with quantum mechanics. The success of DFT is mainly based on its excellent balance between computational cost and accuracy, which is obtained with popular density functionals, such as B3LYP[11-14]. B3LYP has shown good accuracy and transferability for thousands of real world systems studied during the past 15 years and therefore has become the most popular density functional for application studies in the chemical literature. However, even though B3LYP shows remarkable accuracy for many real world problems,[15,16,29,68] it fails to accurately represent London Dispersion Interactions.[17,18,102] Since this type of weak non-covalent interactions plays a crucial role in chemistry,[103-105] biology,[106-108] and nanotechnology,[109] this represents a serious limitation for the B3LYP functional. A variety of ways to improve upon B3LYP for non-covalent interactions have been proposed in the recent literature, ranging from new density functionals[26,28,110] such as M06-2X[27] to *a posteriori* corrections for existing density functionals,[34,35,111,112] to name just a few of the most popular approaches. For a more comprehensive discussion of such methods we refer the reader to the recent review literature.[36,102,113,114]

It becomes an important task to benchmark the significant number of promising DFT methods in order to be able to recommend the best ones to the users. For this it is crucial to have benchmark datasets as large and diverse as possible in hand. The field of producing benchmark interaction energy data has been pioneered by Hobza and coworkers with their S22 dataset[98] which has become a standard test set for non-covalent

interactions. Since the publication of the S22 dataset, a substantial amount of new high level benchmark non-covalent interaction energy data has appeared in the literature.[21] However, since most of this data is scattered throughout the literature, assembling all of it into one database is necessary, if one wants to use it for benchmarking or training purposes. This task has been started by Hobza and coworkers with the Benchmark Energy & Geometry Database (BEGDB) database,[115] which contains most of the benchmark data produced in the Hobza group. In the work described in this chapter I have tried to complete one of the tasks started with the BEGDB project, which is to create one large benchmark interaction energy database that contains almost all CCSD(T) data currently available in the literature. With a thorough literature search I was able to assemble a dataset of 2027 counterpoise corrected interaction energies at the CCSD(T) level. Furthermore, I also compiled a database of CCSD(T)/cbs level conformational energies of di-, tripeptides, sugars and cysteine from the literature.[35,115,116] For future use, our databases are supplied as a whole in the supporting information of reference 21 in their current forms. However, an updated version of both databases will also be made accessible soon on the Friesner group webpage at http://www.noncovalent.friesnerlab.com.

I then employed our new very large and diverse interaction energy database to benchmark two DFT methodologies that both have displayed good performance for non-covalent interactions, M06-2X[27] (a relatively new density functional from the Truhlar group) and B3LYP-D3[34,35] (an *a posteriori* correction scheme for the popular B3LYP density functional developed by Professor Grimme) when evaluated using smaller data sets. While B3LYP-D3 involves an empirical functional form as well as a few global,

empirical parameters, most of its parameters are calculated using *ab initio* methods. Therefore B3LYP-D3 is generally referred to as a *semiempirical* correction scheme.[35] Using part of our new interaction energy and conformational energy dataset as the training set, we also parameterized a novel empirical correction scheme for the B3LYP functional (abbreviated as B3LYP-MM). Our correction scheme not only improves the treatment of London Dispersion Interactions with B3LYP but also includes a version that corrects for basis set superposition error (BSSE)[117] if this is not treated explicitly. It was specifically designed to yield good accuracy with any basis set with and without explicit counterpoise corrections.[118] Since B3LYP-D3 and M06-2X were not designed to correct for the large BSSE present with small basis sets, it is not surprising that we find major improvements in accuracy compared to the other benchmarked methods for the small LACVP*[119] basis set (which is very popularly used because of its relatively low computational cost) without explicit counterpoise corrections. While with B3LYP-MM we get an overall mean unsigned error (MUE) of only 0.41 kcal/mol without explicit counterpoise correction and the LACVP* basis set, the corresponding MUE's with the other methods are significantly larger with 1.20 kcal/mol for M06-2X and 2.11 kcal/mol for B3LYP-D3. With our new correction scheme special care was also taken to obtain high accuracy for hydrogen-bonded systems (for which BSSE is often especially large due to the small distances involved) and for systems with ionic interactions, including those with cation-π interactions.

Our correction is simply added to the B3LYP energy and depends only on nuclear coordinates in the same spirit as the correction schemes developed by Grimme and coworkers.[34,35] It therefore has the big advantage of very low numerical complexity,

which makes implementation of gradients and second derivatives into any existing quantum chemical code a relatively simple task. Furthermore, the time needed to calculate the correction is negligible, making the corrected DFT as fast as the original uncorrected version. The correction consists of three additive parts: A simple Lennard-Jones potential, a linear hydrogen bonding correction term as well as a linear cation-π correction term. Currently, the parameters of the Lennard-Jones correction term depend only on atomic numbers and are independent of the atomic environments, while the hydrogen bonding correction parameters are identical for all types of hydrogen bonds.

Applying combination rules for the pair wise parameters of the Lennard Jones terms allows us to keep the total number of freely adjustable parameters low. Currently, the correction contains only one parameter for each atom type (currently H, C, N, O, F, S and Cl are implemented). For positive ions such as Li^+ or Na^+ Lennard-Jones terms are not needed. Lennard-Jones correction terms are also excluded for ammonium hydrogens (any hydrogen attached to a positively charged sp^3 nitrogen) since such hydrogens carry a substantial positive charge. Besides, there are four additional adjustable parameters, one global scaling factor for Van der Waals radii, two parameters for the hydrogen bonding correction as well as one adjustable parameter for the cation-π correction term. All 11 adjustable parameters were optimized by a non-linear least squares fitting algorithm using part of our very large and diverse dataset of non-covalent interaction energies as well as part of the dataset of conformational energies as the training set.

Our correction scheme is able to yield better accuracy for small basis sets, hydrogen bonded and charged systems mainly due to the following five major differences to similar solutions developed by others:[34-36]

(1) The functional form of our dispersion correction consists of a simple Lennard-Jones 6-12 potential, in contrast to most other correction schemes that contain attractive r^{-6} and r^{-8} terms damped at short distances.

(2) Our correction is only applied for atom-pairs more than three covalent bonds apart from each other, since we believe that special corrections should be developed for the covalent bonding region. Such corrections for covalent bonds as well as corrections for transition metals are currently being developed in the friesner group.[21-25] Our final goal is to merge all different correction types into a general correction scheme for B3LYP.

(3) Hydrogen bonds are treated specially with our correction scheme. First, the Lennard Jones dispersion correction term is not included for any hydrogen heavy atom pair involved in a hydrogen bond. Second, a linear repulsive correction term is applied for all hydrogen heavy atom pairs involved in a hydrogen bond, mainly in order to correct for BSSE (which we find to be especially strong for hydrogen bonds due to the short distances between the atoms involved).

(4) Cation-π interactions obtain a special treatment with our correction scheme as well. First, the Lennard-Jones dispersion correction terms are not included for any positively charged metal ions such as Li^+ or Na^+ or for any ammonium hydrogens (any hydrogen attached to a positively charged sp^3 nitrogens) since the positive charge of an ammonium type cation is mostly localized on the hydrogens. Second, a linear repulsive correction term is applied for cation-π interactions involving simple metal cations such as Li^+ or Na^+.

(5) Since BSSE is highly dependent on the basis set used, the parameters of the correction were made dependent on the basis set and on whether counterpoise corrections are applied.

The final parts of this dissertation chapter are organized as follows. In sections 3.2 – 3.6, our dataset and correction scheme are discussed in detail. Then, in the results section, the performance of our B3LYP-MM correction scheme will be compared to the accuracy of M06-2X and B3LYP-D3. Finally, in the conclusion, we summarize the comparisons of the various approaches.

3.2 Assembly of the largest ever benchmark interaction energy database

I assembled a non-covalent benchmark interaction energy database from 34 publications in the literature.[21] Relevant data from the BEGDB project[115] was downloaded from http://www.begdb.com. Some publications did not report molecular coordinates but I was able to obtain most of them through correspondence with the authors.

The database contains mostly interaction energies at the counterpoise corrected CCSD(T)/cbs level, which represents the current gold standard for benchmark non-covalent interaction energies of medium sized molecules.[120-123] The acronym CCSD(T)/cbs represents MP2[124,125] interaction energies (extrapolated to the complete basis set limit), which have been corrected for higher order correlation effects at the CCSD(T) level[126,127] with a small or medium size basis set.[98,123] Recent studies have shown that for some complexes of the S22 dataset, errors due to small basis sets employed to compute the CCSD(T) correction terms can be as large as 0.6 kcal/mol.[128,129] Unfortunately, up to date, only a few highly accurate CCSD(T)/cbs interactions energies

(with the CCSD(T) correction term extrapolated to the basis set limit) exist due to the extraordinary cost of such calculations, especially when applied to medium sized molecules. Since for this study our main goal was to build a very large and diverse dataset with reasonable accuracy, we assembled our interaction energy dataset from almost all CCSD(T)/cbs interaction energies published in the literature without further pruning of the dataset with respect to quality of the CCSD(T) corrections. The dataset will however be updated in the future to reflect better quality best estimates for interaction energies as they will become available in the literature. Furthermore, the very flexible B3LYP-MM methodology will easily allow refitting of the correction parameters to a dataset of more accurate interaction energies as these become available. Besides CCSD(T)/cbs level interaction energies, several data points with interaction energies at the counterpoise corrected CCSD(T) level obtained with a medium or large basis set are also included in the database. In total, the database contains 2027 counterpoise corrected CCSD(T) interaction energy data points. For the benchmark studies reported in this work, non-covalent complexes with strongly repulsive interaction energies (> 5.0 kcal/mol) were not included, mainly because the precise accuracy of the model at such geometries is generally unimportant in practical applications, as these geometries are rarely occupied due to the strongly repulsive interaction energy. Complexes containing elements other than H, C, N, O, F, S, Cl, Li and Na were not included as well, since currently parameters for B3LYP-MM were only fit for these elements. Even though a few complexes were not used for the benchmark studies, their best estimate interaction energies as well as their coordinates are still given in the supplementary information of reference 21 for potential future use.

The database contains equilibrium structures (96 data points, optimized at the CCSD(T), RI-MP2, MP2 or DFT level) and non-equilibrium structures (1931 data points). It also spans the whole range of non-covalent interactions, ranging from weakly bound dispersion dominated complexes over hydrogen bonded ones to dimers with strong electrostatic interactions. A split up of the database (only the data points that were used for the benchmark studies) into different interaction energy type categories is shown in Table 3.1.

Table 3.1. Non-covalent interaction types present in the interaction energy database (Only data used for the benchmark studies in this work was counted. For all the data see the supplementary information of reference 21).

	Number of Data points	Interaction Energy Range (min/max in kcal/mol)
Dispersion and/or Dipole-Dipole Dominated	1036	-11.83 / 3.71
Hydrogen Bonded, not charged	137	-22.95 / 0.61
With Ionic Interactions	542	-110.80 / 4.62

3.3 A benchmark data set of relative conformational energies

A dataset of conformer energies at the CCSD(T)/cbs level containing di- and tripeptides as well as sugars and cysteine was compiled from the literature as well.[21] The collection of all possible energy gaps between the different conformers (a total of 700) was used as the benchmark set. The peptide benchmark energies were taken from the BEGDB project of Hobza and coworkers,[115] while benchmark data for the sugars and cysteine was taken from the SCONF and the CCONF datasets of Grimme and coworkers.[35] All benchmark conformer energies as well as the corresponding structures

are given together with their original citations in the supplementary information of reference 21.

3.4 General computational Details

All calculations were carried out with the Jaguar software package.[130] The pseudospectral methodology,[50-53] which significantly speeds up the SCF iterations, was employed. Default grids and SCF convergence criteria as implemented in Jaguar were used. All structures were taken from the benchmark energy database and were not further optimized. Interaction Energies were obtained by subtracting the energies of the monomers (in the geometries of the complex) from the energy of the complex. Counterpoise corrected interaction energies were obtained by employing the methodology of Boys and Bernardi.[118] With this methodology, the energies of the monomers are calculated in the basis set of the complex, i.e. basis functions are placed at the positions of all the atoms present in the complex but absent in the monomer. B3LYP-D3[35] dispersion corrections were calculated with the program provided on Professor Grimme's website (at http://toc.uni-muenster.de/DFTD3).

3.5 Near linear dependencies of basis functions

When one tries to solve the SCF equations for non-covalently bound complexes, one is often confronted with the well-known problem of numerical instabilities arising due to near linear dependencies of non-orthogonal gaussian basis functions on atoms close in space.[131] The near linear dependencies are especially pronounced for large basis sets with diffuse functions and can lead to numerical instabilities, especially if approximate numerical techniques (such as the pseudospectral method) are employed to

compute integrals. Usually, the closer the non-covalently bonded atoms, the more pronounced the numerical instabilities. The near linear dependencies are manifested in very small eigenvalues of the atomic orbital overlap matrix. Therefore, the simplest solution to the problem of basis set overcompleteness is to diagonalize the overlap matrix to get a set of *canonical* orbitals as eigenvectors. Then, all *canonical* orbitals with a corresponding eigenvalue below a threshold ε are discarded. This simple methodology effectively removes near linear dependencies of the basis functions and has been successfully applied with Jaguar with $\varepsilon = 5.0\text{x}10^{-4}$.

One caveat with specifying a fixed eigenvalue cutoff ε is that the number and type of canonical orbitals removed can vary with the molecular geometry. Since the total number of canonical orbitals can change abruptly between two infinitesimally close geometries, the potential energy surfaces can become discontinuous. With $\varepsilon = 5.0\text{x}10^{-4}$ these discontinuities are virtually non-existent with the LACVP* basis set. With the aug-cc-pVDZ basis set[132-135] and $\varepsilon = 5.0\text{x}10^{-4}$ the discontinuites are usually small (below 0.1 kcal/mol), they can however in a few cases reach larger values up to 0.3 kcal/mol. We believe that for most applications discontinuities of this magnitude are tolerable, especially since errors from other sources (e.g. stemming from the approximate treatment of electron correlation effects in DFT or from approximate solvation models, etc.) are often of much larger magnitude. However, if very high accuracy is required, a lower cutoff ε might have to be employed in order to minimize the discontinuities further.

Another caveat is that the number and type of canonical orbitals removed can vary between calculations that are used for energy comparisons. For example, since near linear dependencies of basis functions are often present for non-covalent dimers, but absent

from the monomeric structures, the dimer is likely to contain fewer canonical orbitals than both monomers together if a fixed value of ε is employed. This will lead to an increased energy of the dimer relative to the energies of the monomers and therefore lead to a less stable complex. However, if the eigenvalue cutoff ε gets small enough, the degrees of freedom of the removed canonical orbitals are already well represented by other orbitals and their removal does not significantly increase the energy of the molecule. Furthermore, the destabilization effect on a complex due to removal of canonical orbitals is in opposite direction of the BSSE, which is always stabilizing the complex. Therefore, one might expect some error cancellation with the BSSE if a fixed ε is employed. In practice we have found, that with $\varepsilon = 5.0\text{x}10^{-4}$, interaction energies at the non-counterpoise corrected B3LYP/aug-cc-pVDZ level agree within a few tenths of kcal/mol with their counterpoise corrected counterparts. The same was also found for interaction energies at the B3LYP/aug-cc-pVTZ level. With the LACVP* basis set, there are usually no canonical orbital eigenvalues below $\varepsilon = 5.0\text{x}10^{-4}$, such that the BSSE is not reduced with this methodology.

3.6 Corrections for London Dispersion and Basis Set Superposition Error

As many of the numerous correction schemes developed by others,[34-36] our correction is based on an *a posteriori* energy correction, solely dependent on nuclear coordinates. Our correction, which comprises of a London Dispersion correction (E_{LDC}), a hydrogen bonding correction term (E_{HBC}) and a cation-π correction term ($E_{\pi+}$) is simply added to the DFT energy, as described in (2.1).

$$E_{B3LYP-MM} = E_{B3LYP} + E_{LDC} + E_{HBC} + E_{\pi+} \tag{2.1}$$

We found that the errors of B3LYP for hydrogen bonded systems and structures with cation π interactions were very different from the errors observed for dispersion and/or dipole-dipole bound complexes. While B3LYP strongly underestimates interaction energies of most dispersion and dipole-dipole bound complexes, it often overestimates interaction energies of hydrogen-bonded complexes or such with cation π interactions, especially if no explicit counterpoise corrections are applied. Major reasons for this special behavior of hydrogen bonded complexes and systems with cation π interactions are the very strong electrostatic interactions involved, but also the weak covalent nature of the resulting interactions. To account for the special nature of hydrogen bonds and cation π interactions, the London dispersion correction term is replaced with a special hydrogen bond or cation π correction term (described below) for hydrogen bonded heavy atom hydrogen pairs as well as for cation π bonded atom pairs.

The London dispersion correction is described as a sum of Lennard Jones (LJ) functions (2.2), where the sum loops over all atom pairs at least four covalent bonds apart from each other.

$$E_{LDC} = \sum_{i<j} \varepsilon_{ij} [(\frac{r_{ij}^{\min}}{r_{ij}})^{12} - 2(\frac{r_{ij}^{\min}}{r_{ij}})^{6}] \qquad (2.2)$$

The parameters $r_{ij}^{\min}$ represent the minimum distances of the LJ functions, while the parameters ε_{ij} represent the values of the correction terms at $r_{ij}^{\min}$. We found that the simple Lennard Jones 6-12 functional form is well suited to represent dispersion correction terms in the attractive and weakly repulsive region. However, for the strongly repulsive region (with non-bonded atoms jammed into one another), the Lennard Jones

potential is likely to greatly overestimate the van der Waals repulsion. Since most structures with strongly repulsive non-bonded contacts represent in fact highly unstable species which are difficult to accurately model even with correlated wave function methods such as CCSD(T), we have not attempted to accurately model the strongly repulsive region with the current version of B3LYP-MM. However, once a larger amount of reliable interaction energy data becomes available for strongly repulsive van der Waals contacts, it will be a relatively easy task to adjust the repulsive functional form of B3LYP-MM to allow accurate modeling of the repulsive region as well. Since Lennard Jones dispersion correction terms are not included for hydrogen bonds as well as for cation π interactions, the repulsive wall of the LJ 6-12 potential does not pose any problem for close hydrogen bond and cation π contacts.

In order to guarantee transferability of the correction, the total number of empirical parameters should be kept small compared to the number of training data points. Therefore, in order to lower the total amount of empirical parameters, we introduced an empirical combination rule for the atom pair wise ε-parameters (2.3).

$$\varepsilon_{ij} = \varepsilon_i \varepsilon_j \tag{2.3}$$

The parameters ε_i depend only on the atomic number of atom i. This keeps the number of parameters low, but at the same time limits the accuracy of the correction, since atoms in different chemical environments will be treated the same. For this work, ε-parameters were only fit for the most common elements in the dataset (H, C, N, O, F, S and Cl). For metal cations such as Li^+ and Na^+ as well as for ammonium hydrogens (defined as any hydrogen attached to a positively charged sp^3 nitrogen) Lennard-Jones dispersion correction terms are not included since the contribution of dispersion to intermolecular

interactions is usually very small for these positive ions. For molecules involving other elements, the correction is currently undefined. However, once even more benchmark interaction energy data becomes available, we are planning to determine ε parameters for more elements.

The pair wise $r_{ij}^{\min}$ parameters are calculated with the additive combination rule shown in (2.4).

$$r_{ij}^{\min} = q(R_i^{VDW} + R_j^{VDW}) \tag{2.4}$$

R_i^{VDW} are experimental Van der Waals Radii obtained from Bondi's compilation[136] (again only dependent on the atomic number of atom i, shown in table 2), while q is a global scaling factor.

Table 3.2. The parameters R_i^{VDW} (obtained from Bondi's compilation[136]).

Element	R^{VDW} (Å)
H	1.20
C	1.70
N	1.55
O	1.52
F	1.47
S	1.80
Cl	1.75

The hydrogen bonding correction term is simply composed of a linear repulsive function with two global parameters r_0^{hb} and b^{hb} (2.5).

$$E_{HBC} = \begin{cases} -\sum_{i<j} b^{hb}(r_{ij} - r_0^{hb}) & \text{if } r_{ij} \le r_0^{hb} \\ 0 & \text{if } r_{ij} > r_0^{hb} \end{cases} \tag{2.5}$$

The sum in equation 5 runs over all hydrogen – heavy atom pairs, which are hydrogen bonded. r_{ij} are the distances between the hydrogen bond acceptor and the hydrogen-bonded hydrogens. Each correction term is only counted if positive, i.e. if $r_{ij} \leq r_0^{hb}$. Hydrogen bonds were detected with the hydrogen bond detection subroutine implemented in the Schrödinger MMShare software package. A hydrogen bond was always assigned if two potentially hydrogen-bonding atoms (hydrogen and hydrogen-bond acceptor with the right Macromodel atom types) were closer than 3.0 Å apart from each other. In order to avoid large discontinuities in the potential energy surfaces with our correction scheme, hydrogen bond donor and acceptor angles were not considered for the assignment of hydrogen bonds. This angle independent hydrogen bond assignment ensures that the only possible discontinuity in the potential energy surface could occur at the cutoff distance of 3.0 Å. However, since all correction terms become negligible at the cutoff distance, this very small discontinuity of the potential energy surface does not pose a problem in practice. This fact is illustrated in Figure 3.1, which shows potential energy surfaces of formic acid dimer with our correction scheme. Even though formic acid dimer contains two symmetric hydrogen bonds (which will lead to a discontinuity twice as large as with only a single hydrogen bond), Figure 3.1 shows that the discontinuity in the potential energy surface is virtually non-existent.

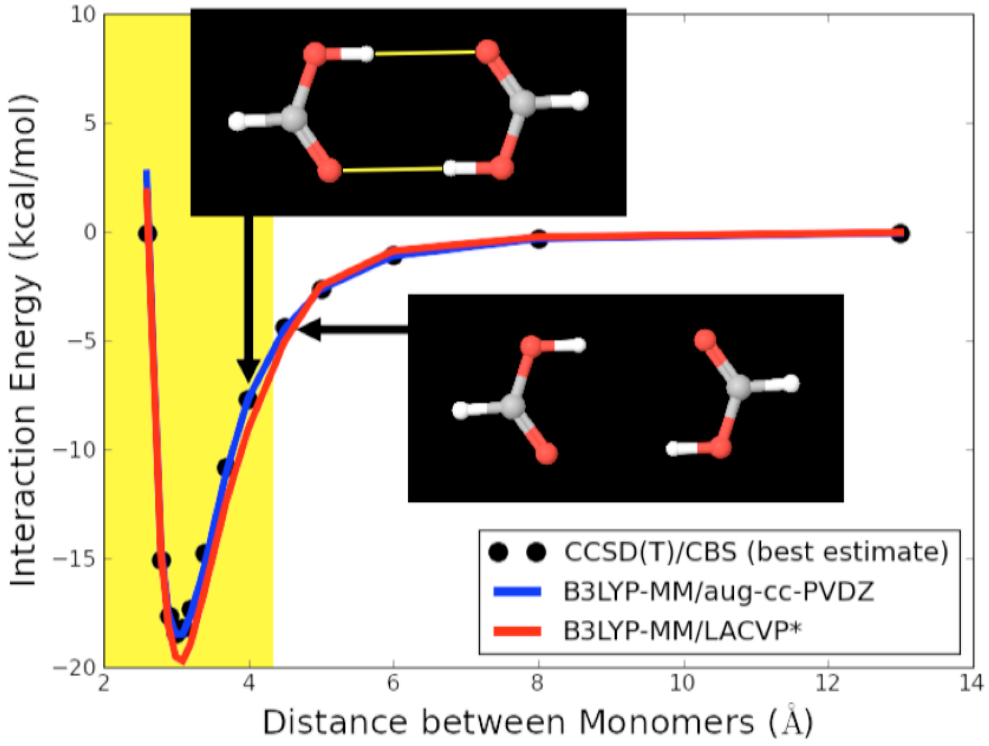


Figure 3.1. B3LYP-MM potential energy surface of the formic acid dimer. Distance between the monomers represents the distance between the two centers of mass. Hydrogen bonds are shown in yellow with the hydrogen bonded distance range (H-heavy-atom distance < 3.0 Å) highlighted in yellow.

The cation-π correction term (2.6) is also composed of a simple linear repulsive term.

$$E_{\pi+} = \begin{cases} -\sum_{i<j} b^{\pi+}(r_{ij} - r_0^{\pi+}) & \text{if } r_{ij} \le r_0^{\pi+} \\ 0 & \text{if } r_{ij} > r_0^{\pi+} \end{cases} \tag{2.6}$$

The sum in (2.6) runs over all metal cation – sp^2/sp-carbon pairs. r_{ij} are the distances between the atom pairs considered. All sp^2/sp-carbons, except carbonyl and immine carbons are counted for the cation-π correction term. Each correction term is only included if positive, i.e. if $r_{ij} \le r_0^{\pi+}$. The parameter $r_0^{\pi+}$ was set to 5.0 Å and not further optimized.

The parameters ε_i, q, b^{hb}, r_0^{hb} and $b^{\pi+}$ were optimized via a least squares fitting procedure to part of our interaction energy data set as well as to part of the data from the relative conformational energy data set. In order to avoid any discontinuity of the hydrogen bonding correction, the hydrogen bonding correction parameter b^{hb} was constrained to be ≤ 3.0 Å. The determination of optimal parameter values represents a non-linear optimization problem, to which we found a solution with a variant of the Levenberg-Marquardt algorithm as implemented in the leastsq subroutine of the scipy software package (version 0.6.0). About 75% of all interaction energy and conformational energy data was randomly selected and used as the training set for the parameters. The leftover 25% of the data was employed as a test set. In order to ensure, that all types of interactions were well represented in the training and the test set, the random selection procedure for the interaction energy training set was carried out according to the following protocol:

(1) All complexes were assigned to several *ε-parameter groups*, such that each parameter ε_i formed one *parameter group*. A complex was assigned to the parameter group of ε_i if its interaction energy was dependent on the value of ε_i. *ε-parameter groups* of common elements (e.g. C, H or O) contained several hundreds of complexes while other *parameter* groups, such as e.g. the ones involving F or Cl only had a few tens of members.

(2) In order to ensure that all *parameter groups* would be represented in the training and the test set with approximately the desired training/test ratio, all *parameter groups* were ordered according to increasing number of complexes. Each complex was then assigned to one single *parameter group* according to the following algorithm: First, the *paramter*

group with the least amount of members was assigned all of its members. Then, the second smallest *parameter group* was assigned all of its members, except the ones already present in the first parameter group. Proceeding in the same manner, the n^{th} smallest parameter group was assigned all of its members except the ones already present in the parameter groups with fewer members.

(3) A little over half of the training set was then selected by randomly choosing 37.5% (rounded down to the nearest integer) of complexes from each *parameter group*.

(4) The whole dataset was also divided into the three *interaction type categories* shown in Table 3.1. The three categories are: complexes with only dispersion and dipole-dipole interactions, hydrogen bonded complexes and charged complexes.

(5) The percentage of complexes to be selected for the training set from each *interaction type category* (defined as *f2*), was then calculated, such that the final training/test ratio would be around 3/1. *f2%* of data points from each *interaction type category* were then randomly selected for the training set, after having removed all complexes already selected for the training set in step 3.

The random selection of the training set together with the least squares fitting was carried out six times for each set of parameters. This allowed us to obtain an estimate of the sensitivity of the parameters on the training set and is a good test to ensure no overfitting. Table 3.3 shows the mean values of all parameters (for the LACVP* and the aug-cc-pVDZ basis sets) together with their standard deviations as obtained from the six least squares fitting experiments. Optimal parameters were determined for counterpoise and non-counterpoise corrected interaction energies. (The relative conformational energies

were never counterpoise corrected, since there is no consistent way to apply counterpoise correction for total energies.

Table 3.3. B3LYP-MM parameter values. All values represent mean values obtained from six non-linear least squares fitting experiments. Error bars represent standard deviations.

Parameter name	Parameter Units	Basis Set			
		LACVP*		aug-cc-pVDZ	
		un-Cp-corr[a]	*Cp-corr*[b]	*un-Cp-corr*[a]	*Cp-corr*[b]
ε_H	$(\text{kcal/mol})^{0.5}$	0.097 ± 0.009	0.183 ± 0.002	0.306 ± 0.003	0.313 ± 0.002
ε_C		0.589 ± 0.024	0.744 ± 0.003	0.660 ± 0.009	0.714 ± 0.013
ε_N		0.542 ± 0.019	0.744 ± 0.005	0.731 ± 0.009	0.705 ± 0.011
ε_O		0.215 ± 0.004	0.427 ± 0.004	0.595 ± 0.012	0.633 ± 0.013
ε_F		0.013 ± 0.023[c]	0.528 ± 0.022	0.362 ± 0.010	0.540 ± 0.014
ε_S		1.117 ± 0.067	1.393 ± 0.023	1.288 ± 0.020	1.379 ± 0.027
ε_{Cl}		0.909 ± 0.043	1.145 ± 0.023	0.701 ± 0.054	0.974 ± 0.044
q	no units	0.895 ± 0.012	0.860 ± 0.001	0.859 ± 0.004	0.846 ± 0.005
b^{hb}	kcal/(mol x Å)	1.144 ± 0.024	1.094 ± 0.038	1.888 ± 0.064	1.816 ± 0.090
$b^{\pi+}$		0.410 ± 0.004	0.248 ± 0.006	0.130 ± 0.002	0.116 ± 0.004
r_0^{hb}	Å	3.000[d]	2.283 ± 0.006	2.047 ± 0.002	2.035 ± 0.006
$r_0^{\pi+}$		5.000[e]	5.000[e]	5.000[e]	5.000[e]

[a] Parameters fit to interaction energies without explicit counterpoise corrections. [b] Fit to counterpoise corrected interaction energies. Relative Conformational energies were never counterpoise corrected. [c] Parameter was constrained to be positive. During the six fitting experiments parameter values from 0.00 to 0.06 were found. [d] Parameter reached its maximum value of 3.0 Å during least squares optimization.[e] Parameter was not optimized but set to 5.0 Å.

The very small standard deviations for each parameter clearly show that there has been no overfitting and that the correction is therefore likely to be transferable to systems not included in the training set. Another fact that supports this conclusion is that the Mean Unsigned Errors (MUEs) as well as the Root mean square deviations (RMSDs) for the randomly selected training and test sets were very similar for all six fitting experiments. RMSDs and MUEs averaged over all six fitting experiments are shown in Table 3.4.

Table 3.4. RMSDs and MUEs obtained with B3LYP-MM averaged over the six fitting experiments. Error bars represent standard deviations. (Training set: 1810 ± 6 data points; Test set: 605 ± 6 data points).

Basis Set	*cp corr* [a]	**MUE (kcal/mol)**		**RMSD(kcal/mol)**	
		Training set	Test set	Training set	Test set
LACVP*	*no*	0.45 ± 0.01	0.45 ± 0.02	0.70 ± 0.02	0.66 ± 0.05
	yes	0.48 ± 0.01	0.49 ± 0.01	0.72 ± 0.01	0.72 ± 0.03
aug-cc-pVDZ	*no*	0.37 ± 0.01	0.36 ± 0.01	0.51 ± 0.01	0.50 ± 0.02
	yes	0.34 ± 0.01	0.35 ± 0.01	0.49 ± 0.01	0.49 ± 0.02

[a] Whether interaction energies were counterpoise corrected or not. Relative Conformational energies were never counterpoise corrected.

Since we find that BSSE is small with the aug-cc-pVDZ basis set if no halogens are involved, it is not surprising that the optimal values for the correction parameters are very similar with and without counterpoise corrected interaction energies for that basis set (Table 3.3). An exception are the halogen correction parameters ε_F and ε_{Cl}, which vary significantly depending on whether explicit counterpoise corrections are included, due to larger counterpoise corrections for complexes involving F or Cl with the aug-cc-pVDZ basis set. For the small LACVP* basis set where BSSE is very significant (often up to several kcal/mol) we find clearly different correction parameters depending on whether counterpoise corrections are applied or not. Again, the halogen correction parameters are especially sensitive to the counterpoise correction issue, due to especially large counterpoise corrections for structures involving halogens.

3.7 Results

A summary of the overall statistical performance (MUEs, RMSDs and Average Errors) of all the benchmarked methods, B3LYP-D3, M06-2X and B3LYP-MM is shown in Table 3.5.

Table 3.5. RMSDs, MUEs and Average Errors of all benchmarked DFT methods (in kcal/mol). B3LYP-MM represents results obtained with the correction scheme developed in this work.

Basis Set	Counter-poise corrected	DFT-Method	Whole Interaction Energy DB (1715 datapoints[a])			Conformational Energies DB (700 data points)		
			MUE[b]	RMSD[c]	Avg[d]	MUE[b]	RMSD[c]	Avg[d]
LACVP*	no	B3LYP-D3	2.11	2.87	-2.10	0.82	1.10	0.02
		B3LYP-MM	**0.41**	**0.68**	**-0.04**	**0.55**	**0.71**	**0.05**
		M06-2X	1.20	1.97	-0.94	0.78	1.04	0.10
		B3LYP	2.75	3.66	1.35	1.54	2.03	0.32
	yes	B3LYP-D3	1.21	1.91	-0.64		na[e]	
		B3LYP-MM	0.41	0.65	0.18	*0.67*[f]	*0.87*[f]	*-0.03*[f]
		M06-2X	1.20	1.97	-0.94		na[e]	
		B3LYP	3.53	4.87	2.81			
aug-cc-pVDZ	no	B3LYP-D3	0.95	1.48	-0.72	0.39	0.49	-0.05
		B3LYP-MM	**0.37**	**0.53**	**0.09**	**0.37**	**0.47**	**-0.07**
		M06-2X	0.73	1.17	-0.30	0.61	0.79	-0.02
		B3LYP	3.11	4.36	2.73	1.57	2.15	0.25
	yes	B3LYP-D3	0.87	1.41	-0.56		na[e]	
		B3LYP-MM	0.32	0.48	0.11	*0.38*[f]	*0.49*[f]	*-0.09*[f]
		M06-2X	0.67	1.08	-0.10		na[e]	
		B3LYP	3.25	4.57	2.89			

[a] For the following levels of theory fewer number of cases were included due to SCF convergence problems: B3LYP/aug-cc-pVDZ without counterpoise correction (2 cases missing); M06-2X/LACVP* with explicit counterpoise correction (11 cases missing); M06-2X/aug-cc-pVDZ with explicit counterpoise correction (3 cases missing); M06-2X/aug-cc-pVDZ without counterpoise correction (4 cases missing) [b] Mean Unsigned Error. [c] Root Mean Square Deviation. [d] Average Error. [e] No consistent way to apply counterpoise corrections. [f] B3LYP energies were obtained without counterpoise corrections, however, the MM correction part was parametrized with counterpoise corrected interaction energies.

In order to see how the benchmarked methods perform for complexes dominated by different kinds of interactions, results for dispersion and dipole-dipole dominated complexes are shown separately in Table 3.6, while Table 3.7 shows separate results for complexes with hydrogen bonds or ionic interactions.

Table 3.6. RMSDs, MUEs and Average Errors of all benchmarked DFT methods for dispersion and dipole-dipole dominated complexes (in kcal/mol). B3LYP-MM represents results obtained with the correction scheme developed in this work.

Basis Set	Counter-poise corrected	DFT-Method	Dispersion and Dipole-Dipole Dominated Complexes (1036 datapoints[d])		
			MUE[a]	RMSD[b]	Avg[c]
LACVP*	no	B3LYP-D3	1.16	1.38	-1.15
		B3LYP-MM	**0.28**	**0.41**	**0.05**
		M06-2X	0.65	0.92	-0.26
		B3LYP	3.15	4.09	3.14
	yes	B3LYP-D3	0.55	0.80	0.33
		B3LYP-MM	0.33	0.46	0.21
		M06-2X	1.06	1.36	1.03
		B3LYP	4.62	5.88	4.62
aug-cc-pVDZ	no	B3LYP-D3	0.37	0.50	-0.02
		B3LYP-MM	**0.36**	**0.49**	**0.13**
		M06-2X	0.53	0.73	0.01
		B3LYP	4.27	5.31	4.27
	yes	B3LYP-D3	0.32	0.48	0.18
		B3LYP-MM	0.27	0.37	0.14
		M06-2X	0.47	0.63	0.31
		B3LYP	4.47	5.58	4.47

[a] Mean Unsigned Error. [b] Root Mean Square Deviation. [c] Average Error. [d] For the following levels of theory fewer number of cases were included due to SCF convergence problems: B3LYP/aug-cc-pVDZ without counterpoise correction (2 cases missing); M06-2X/aug-cc-pVDZ without counterpoise correction (4 cases missing).

Table 3.7. RMSDs, MUEs and Average Errors of all benchmarked DFT methods for complexes with hydrogen bonds or ionic interactions (in kcal/mol). B3LYP-MM represents results obtained with the correction scheme developed in this work.

Basis Set	Counter-poise corrected	DFT-Method	Hydrogen Bonded Complexes (not charged) (137 data points)			Complexes With Ionic Interactions (542 datapoints[d])		
			MUE[a]	RMSD[b]	Avg[c]	MUE[a]	RMSD[b]	Avg[c]
LACVP*	no	B3LYP-D3	3.42	3.71	-3.42	3.58	4.36	-3.57
		B3LYP-MM	**0.75**	**0.94**	**-0.31**	**0.56**	**0.95**	**-0.15**
		M06-2X	1.85	2.16	-1.81	2.08	3.07	-2.03
		B3LYP	2.13	2.76	0.14	2.14	2.93	-1.76
	yes	B3LYP-D3	0.82	1.00	-0.51	2.57	3.18	-2.53
		B3LYP-MM	0.55	0.91	0.34	0.52	0.84	0.10
		M06-2X	0.85	1.01	0.53	1.35	2.11	-1.16
		B3LYP	3.13	4.21	3.04	1.53	2.13	-0.71
aug-cc-pVDZ	no	B3LYP-D3	0.64	0.77	-0.45	2.14	2.51	-2.12
		B3LYP-MM	**0.37**	**0.54**	**0.10**	**0.38**	**0.58**	**0.00**
		M06-2X	0.61	0.76	0.25	1.13	1.78	-1.05
		B3LYP	3.10	4.05	3.10	0.91	1.47	-0.31
	yes	B3LYP-D3	0.44	0.56	-0.25	2.05	2.41	-2.04
		B3LYP-MM	0.36	0.51	0.24	0.41	0.64	0.02
		M06-2X	0.63	0.78	0.52	1.09	1.67	-1.05
		B3LYP	3.31	4.20	3.31	0.90	1.44	-0.23

[a] Mean Unsigned Error. [b] Root Mean Square Deviation. [c] Average Error. [d] For the following levels of theory fewer number of cases were included due to SCF convergence problems: M06-2X/LACVP* with explicit counterpoise correction (11 cases missing); M06-2X/aug-cc-pVDZ with explicit counterpoise correction (3 cases missing).

Error histograms for dispersion and dipole dominated complexes, but also for hydrogen bonded and ionic ones are shown in Table 3.8 for B3LYP-MM, B3LYP-D3 and M06-2X. Corresponding error histograms are also shown for the conformational energy dataset. Non-counterpoise corrected results are depicted for both basis sets. In the following section we will discuss the results obtained with both basis sets in more detail.

3.8 Discussion of the medium size basis set results

From Table 3.6 it can be seen that the performance of B3LYP-D3 for dispersion and dipole-dipole dominated complexes is remarkably accurate with the aug-cc-pVDZ basis set (MUEs below 0.40 kcal/mol). B3LYP-D3 also performs very accurately for the

dataset of conformational energies (MUE of 0.39 kcal/mol). M06-2X gives also rather accurate results for dispersion and dipole-dipole dominated complexes as well as for the conformational energies. Its performance is however slightly worse than B3LYP-D3's. For the subset of hydrogen bonded complexes B3LYP-D3 is slightly more accurate than M06-2X with explicit counterpoise corrections, while without counterpoise corrections both methods are of comparable accuracy. For hydrogen-bonded systems, B3LYP-D3 shows some overbinding. This overbinding is present with and without counterpoise corrections and is manifested in the negative average errors of the method. Since BSSE is always attractive, it is not surprising, that the overbinding is more pronounced without explicit counterpoise corrections. For complexes with ionic interactions, both B3LYP-D3 and M06-2X show significant overbinding regardless of counterpoise corrections (average errors < -1.0 kcal/mol).

With the aug-cc-pVDZ basis set and with explicit counterpoise corrections, our B3LYP-MM method is slightly more accurate than B3LYP-D3 for dispersion and dipole-dipole dominated complexes (MUE = 0.27 kcal/mol). However, without counterpoise corrections, our B3LYP-MM method still produces very accurate interaction energies. As with B3LYP-D3 we also observe a very accurate performance on the dataset of conformational energies (MUE = 0.37 kcal/mol). However, in contrast to B3LYP-D3, which shows significant overbinding for hydrogen-bonded and charged complexes, the B3LYP-MM correction scheme also delivers results of high accuracy (MUEs ≤ 0.41 kcal/mol) and almost no overbinding (|Average Error| ≤ 0.24 kcal/mol) for hydrogen-bonded and charged systems.

Therefore, since hydrogen bonds as well as ionic interactions are of great practical importance in biology and materials science, B3LYP-MM clearly displays a practical advantage over B3LYP-D3 or M06-2X in conjuction with a medium size basis set.

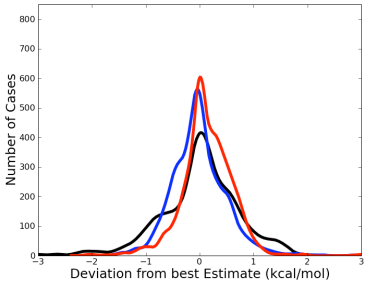
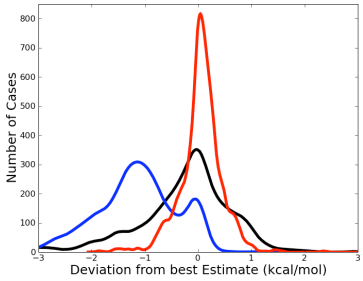
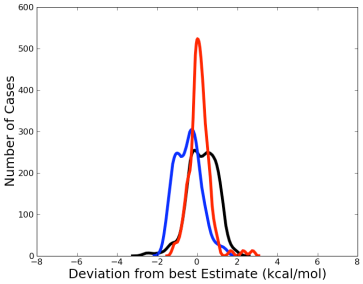
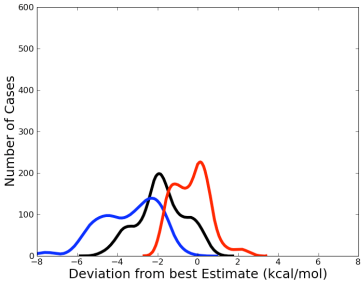
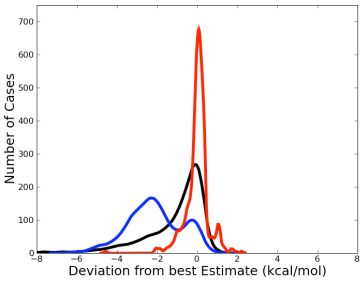
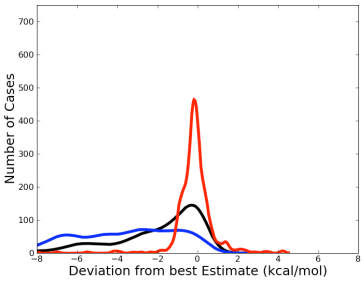
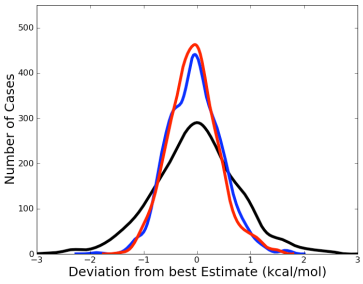
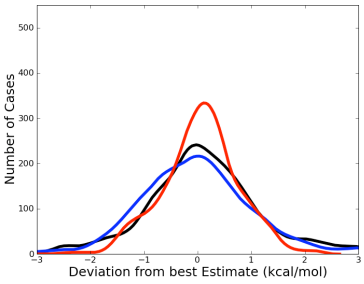
3.9 Discussion of results obtained with the small LACVP basis set*

With counterpoise corrections, B3LYP-D3 performs acceptably in the small basis set for dispersion and dipole-dipole dominated complexes, with an overall MUE of 0.55 kcal/mol, while we would not recommend M06-2X with this basis set (MUE = 1.06 kcal/mol). However, B3LYP-MM is clearly the most accurate method with counterpoise corrections in the small basis set with an overall MUE of only 0.33 kcal/mol. Even with counterpoise corrections, B3LYP-D3 seriously overbinds hydrogen-bonded complexes (Average Error = -0.51 kcal/mol) and complexes with ionic interactions (Average Error = -2.53). Without counterpoise corrections, B3LYP-D3's overbinding for hydrogen-bonded and ionic complexes clearly renders the method very inaccurate with average errors < -3.0 kcal/mol. Apart from hydrogen bonds and ionic interactions, B3LYP-D3 also shows serious overbinding for dispersion and dipole-dipole dominated complexes without counterpoise corrections in the small basis set as can clearly be seen from the error histograms in Table 3.8. Without counterpoise corrections, M06-2X performs better than B3LYP-D3 in the small basis set, its overall performance is however still not convincing with a MUE of 1.20 kcal/mol.

In contrast to all other methods tested, our B3LYP-MM methodology still delivers accurate interaction and conformational energies with the small basis set without counterpoise corrections, with overall MUEs of only 0.41 and 0.55 kcal/mol that are only marginally larger than the ones obtained with the medium size basis set. For hydrogen-

bonded complexes, B3LYP-MM shows slight overbinding (Average Error of -0.31 kcal/mol), which leads to a slight degradation in the overall accuracy for hydrogen-bonded complexes (MUE = 0.75 kcal/mol). However, compared to the performances of B3LYP-D3 and M06-2X for hydrogen-bonded complexes without counterpoise corrections in the small basis set (both methods show MUEs larger than 1.8 kcal/mol), the accuracy of B3LYP-MM represents a significant improvement. For complexes with ionic interactions, B3LYP-MM gives also very accurate results regardless of explicit counterpoise corrections, with MUEs $\leq$ 0.56 kcal/mol.

Table 3.8. Comparison of Error Distributions for non-covalent interactions obtained with B3LYP-MM (red), B3LYP-D3 (blue) and M06-2X (black). All results are without explicit counterpoise corrections. All plots show normalized kernel density estimates (kdes) multiplied with the number of datapoints. The kdes were computed with the density() function in R, with gaussian kernels and default bandwidths.

	aug-cc-pVDZ basis set	LACVP* basis set
Dispersion and Dipole-Dipole Dominated Complexes		
Hydroge Bonded Complexes (not charged)		
Complexes with ionic interactions		
Relative Conformational Energies		

3.10 Further assessment of the cation π correction in B3LYP-MM

It is worth to mention that the cation π correction in B3LYP-MM seems to pretty uniformly improve cation π interaction energies, regardless of whether the cation sits on top or on the side of an aromatic ring. To illustrate this, we calculated the MUE's obtained with B3LYP-MM for all benzene alkalication interaction energies in two different regimes of the angle θ (defined in Figure 3.2).

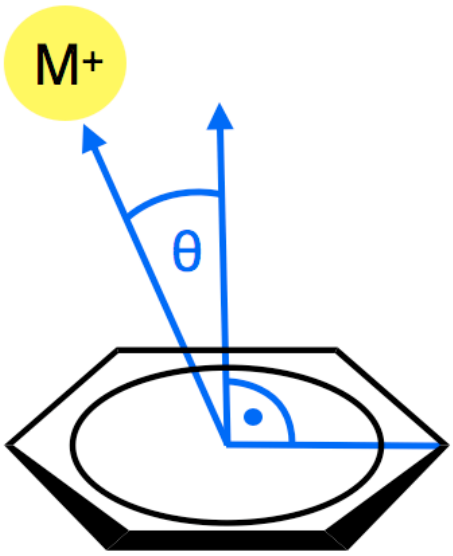


Figure 3.2. Definition of the angle θ used to further assess the accuracy of the cation π correction term in B3LYP-MM.

All structures were divided into conformations, where the cation sits on top of the benzene ring ($\theta \leq 45°$; 215 structures) and in conformations, where the cation is located more towards the side of the benzene ring ($45° < \theta \leq 90°$; 293 structures). With B3LYP-MM we find similar accuracy for both regimes of the angle θ regardless of basis set and

explicit counterpoise corrections. The observed MUEs for the regime with $\theta > 45°$ are in fact always slightly smaller (maximum difference in MUEs observed was 0.27 kcal/mol) than the corresponding MUEs for the regime with $\theta \leq 45°$. This indicates that the cation π correction contained in B3LYP-MM not only works well if the cation is directly above the benzene ring, but also clearly improves interaction energies for other structures.

3.11 Concluding remarks about our correction scheme

I have built a dataset of highly accurate non-covalent interaction energies at the CCSD(T) level from the literature. Compared to previously published datasets, my dataset has a much larger size and greater diversity. Benchmark studies of two DFT methodologies (M06-2X and B3LYP-D3) on my new interaction energy dataset show high accuracy of B3LYP-D3 for dispersion and dipole-dipole dominated interactions if a medium size basis sets such as aug-cc-pVDZ is employed with explicit counterpoise corrections (MUE = 0.32 kcal/mol). However, without counterpoise corrections, serious overbinding is observed with B3LYP-D3 for hydrogen-bonded systems (Average Error = -0.45 kcal/mol) and for complexes with ionic interactions (Average Error = -2.12 kcal/mol) with the medium size basis set. For the small LACVP* basis sets without counterpoise corrections (a level of theory that is often employed in practice, especially for geometry optimizations), our benchmark results show that neither B3LYP-D3 nor M06-2X shows good accuracy. In contrast to B3LYP-D3 and M06-2X, we find that our new methodology (B3LYP-MM) delivers very accurate results for all types of interactions, regardless of counterpoise corrections and basis set (Overall MUEs $\leq$ 0.41 kcal/mol). The observed improvement in accuracy with B3LYP-MM over B3LYP-D3 and M06-2X is especially significant for hydrogen-bonded and charged systems.

The development of B3LYP-MM has demonstrated a number of interesting points:

(1) By making parameters specific to basis set and use of counterpoise corrections, significant improvements in accuracy can be obtained, with a particularly large improvement for small basis sets.

(2) The addition of a small number of parameters addressing hydrogen bonding and interactions involving one or more ions can substantially improve accuracy for these systems, with a particularly large impact on ionic systems, which exhibit large errors when treated with prior alternatives in the literature.

While the data set here has quite a few data points, it does not fully cover all possible types of chemistries associated with noncovalent interactions. Unusual functional groups may in fact require the addition of new parameters to the model. The improvements to hydrogen bonded and charged systems obtained by following this path suggests that it is a fruitful one and can result in a systematically improvable, empirically corrected DFT functional; as problems are discovered, they can readily be repaired by the addition of a relatively small number of new parameters. As benchmark quantum chemical calculations become increasingly inexpensive in the future, transfer of information from such calculations to DFT-based models will become more facile, and, as shown here, will be a highly effective means of achieving chemical accuracy at modest computational cost.

As already mentioned above, our new correction scheme not only gives accurate results with large, but also with faster, small basis sets in contrast to the correction schemes developed by Grimme and coworkers. This will likely be of particular value when applied to predict enantioselectivities, as geometry optimizations with large basis

sets are too expensive for such predictions and single point calculations with large basis sets were shown to give unsatisfactory results on our epoxidation dataset (see section 2.3). Therefore, the next step towards a more accurate prediction protocol for enantioselectivities is to apply our corrections to the prediction methodology described in chapter 2. Recent progress towards this goal will be summarized in the following chapter.

4 Improvents to the enantioselectivity prediction methodology

4.1 Automation of the prediction protocol

All transition state optimizations of the original test set of enantioselective dioxirane catalyzed epoxidations were set up manually by me and Michelle Hall, which involved a painful amount of repetitive work. While such an approach was well suited to carry out an initial assessment of how accurately current density functional theory could predict enantioselectivities, it would involve too much human work to generalize such a manual approach to larger datasets containing different types of enantioselective reactions. Therefore, in order to make our prediction methodology user-friendly, I have now completely automated the setup of all transition state calculations. The user interface of the current approach is based on a Django website, which interacts with a powerful SQL database to store data input by the user as well as the results of the calculations. The website uses the MarvinSketch and the MarvinView java applets from ChemAxon, which provide a ChemDraw like feel for the user to input structures and look at the results. The user now only has to choose the reaction type and then draw all products and the catalyst of the reaction to start a calculation. The program will then automatically setup initial guesses for all relevant transition states, starting materials and products involved in the reaction and carry out all calculations needed to do the selectivity predictions.

The main reason for choosing a procedure where the user inputs all products rather than the starting materials of the reactions is based on the generality of the resulting methodology. If the user draws all products he or she is interested in, the program can simply output the calculated selectivities associated with each of the

products input by the user, regardless of how many and what type of products there are. For now, these products are simply enantiomers, however, once we start to predict reactions where diastereo- or even regioselectivities also play a role, our current methodology will be easy to generalize to handle such situations as well. On the contrary, if the user would just input the starting materials, there would be no way for the program to know in which of the products the user is actually interested in and it would need to consider all possible reaction products, even if the user might only show interest in a particular type of selectivity.

With a completely automatic prediction protocol in hand, we are now working on benchmarking our prediction methodology for new reactions, one of them being the popularly used CBS reduction.[37] In order to get sufficient conformational sampling of the transition state structures for such reactions with more flexible transition states, I also improved the way in which the conformational search of the transition states is carried out. Our new protocol for doing the transition state conformational search will briefly be summarized in the following section.

4.2 An improved conformational searching algorithm

Hand in hand with the complete automation of the prediction methodology, I also improved the way, by which the initial conformational searching of the transition state structures is carried out. In our original protocol (described in chapter 2), the conformational search of the transition states was simply carried out by keeping the cartesian coordinates of all key transition state atoms frozen. While this simple approach worked well for localized and relatively rigid transition states such as the ones of

dioxirane epoxidations, it might not be able to find good guesses for the lowest energy transition state conformations in the case of more flexible transition states. Since many important enantioselective reactions such as CBS reductions or amino acid catalyzed asymmetric aldol condensations do in fact have flexible cyclic transition states,[137-139] I decided to generalize the conformational search protocol, such that it will also be able to handle such transition states well.

Our improved conformational search methodology employs a very simple transition state force field, which pulls the transition state atoms into transition state like arrangements upon minimization and conformational search with MacroModel. The transition state force field uses bonding, angular and dihedral torsion terms as defined in the OPLS2005 force field, parameters of which were simply adjusted to accurately reproduce the key geometric features of model transition states. The simplicity of the transition state force field makes it straightforward to generalize it to new reactions. Work in this direction is currently under way as well.

While this quick summary of our still ongoing improvements to our enantioselectivity prediction protocol is by no means conclusive, detailed results obtained with our improved protocol as well as computational details will be reported in due time in the literature.

5 Structure based predictions of P450 metabolites with IDSite

5.1 IDSite, a structure and physics based methodology

As already introduced in chapter 1, prediction of metabolic reactions catalyzed by P450 enzymes is of great importance for drug discovery, since accurate prediction of the metabolic products helps experimental chemists to focus their synthetic efforts towards drug candidates with desirable metabolic properties. In order to estimate the major metabolic reaction products, the relative transition state free energies of the rate determining reaction step have to be estimated, which remains a very challenging problem. One major difficulty is the high conformational flexibility of the enzymatic active site, which shows very significant induced-fit effects, especially with larger ligands. In order to model such induced-fit effects a large amount of conformational sampling is needed, which can not easily be carried out with QM/MM methods.[41] Therefore, force field based methods, which are several orders of magnitude faster than QM/MM to evaluate the energy of a system, are currently the practical solution for making general predictions for ligand of practical sizes for drug discovery. However, the limits in energetic accuracy of current forcefields also make accurate predictions of enzymatic reactivity a very challenging problem.

Despite all of the above-mentioned great challenges, Dr. Jianing Li and I have developed an accurate structure- and physics-based methodology to predict P450 SOMs. Our protocol, IDSite[46], uses fast forcefield based methods in combination with efficient sampling and a physical scoring model to evaluate the structures and relative free energies of all transition-state-like structures. The scoring model simply depends on the forcefield energies of those structures as well as the intrinsic reactivities estimated from

quantum mechanical DFT calculations on model systems. To estimate the transition state structures, physical constraints (obtained from DFT calculations on model systems as well) are applied to the transition state regions to pull the ligands into transition state like conformations. Additional details about our new IDSite methodology for predicting P450 SOMs are available in the dissertation of Dr. Jianing Li[45] as well as in reference 46. While Dr. Jianing Li has been the leader for the development of the force field based sampling methods, my main contribution to the project has been the design of the intrinsic reactivity model.

Our IDSite prediction protocol was initially tested on a diverse set of 56 P450 2D6 substrates[140] and showed outstanding accuracy for this set of compounds, with very low false positive *and* false negative rates. These very low false positive and false negative rates represent a very significant improvement over previous studies, which reported much larger false positive rates at similar true positive rates.[141] Furthermore, IDSite also produced reasonable guesses for the rate determining P450 metabolism transition states for all of these substrates, which most likely include the important induced-fit effects. These highly accurate transition state guesses could be great starting points for even more accurate transition state optimizations with QM/MM[42], which might reveal additional valuable information about the reaction mechanisms.

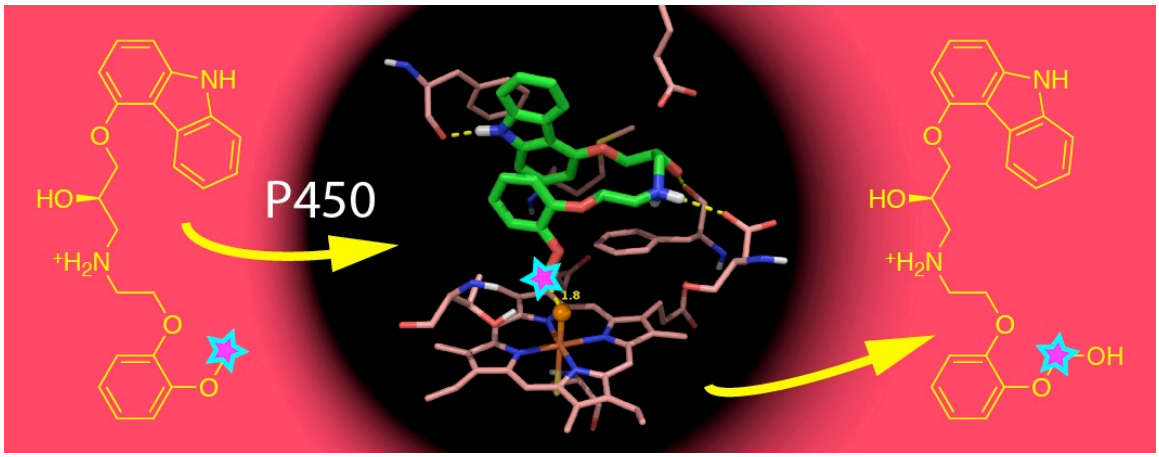


Figure 5.1. A typical metabolic reaction of carvedilol with P450 2D6, leading to a demethylated metabolite. The methylene group, which gets hydroxylated is highlighted with stars. Note the nicely packed structure of the ligand in the reactive conformation, with two polar contacts (yellow, dotted lines). The position of the reactive oxo-atom is shown as an orange sphere above the Heme group.

The great potential of IDSite in modeling induced fit effects of transition state like conformations is illustrated in Figure 5.1, which shows the lowest energy structure for carvedilol leading to a demethylated metabolite, a reaction also observed experimentally.[140] Figure 5.2 compares the ROC curves of the most promising previously published methods[141] for predicting P450 2D6 SOMs to the ROC curves obtained with IDSite.[46] While it is important to mention that our dataset was not the exactly the same as the one used for the other methods[141], it seems clear that our new methodology is significantly more accurate than the previous ones. Last, but not least, IDSite also likely to be much more transferable to other P450 isoforms and maybe even to other enzymatic reactions as it is completely physics based. Studies of such generalizations of our IDSite methodology are currently under way and will be reported in due time. Preliminary initial results look very promising however.

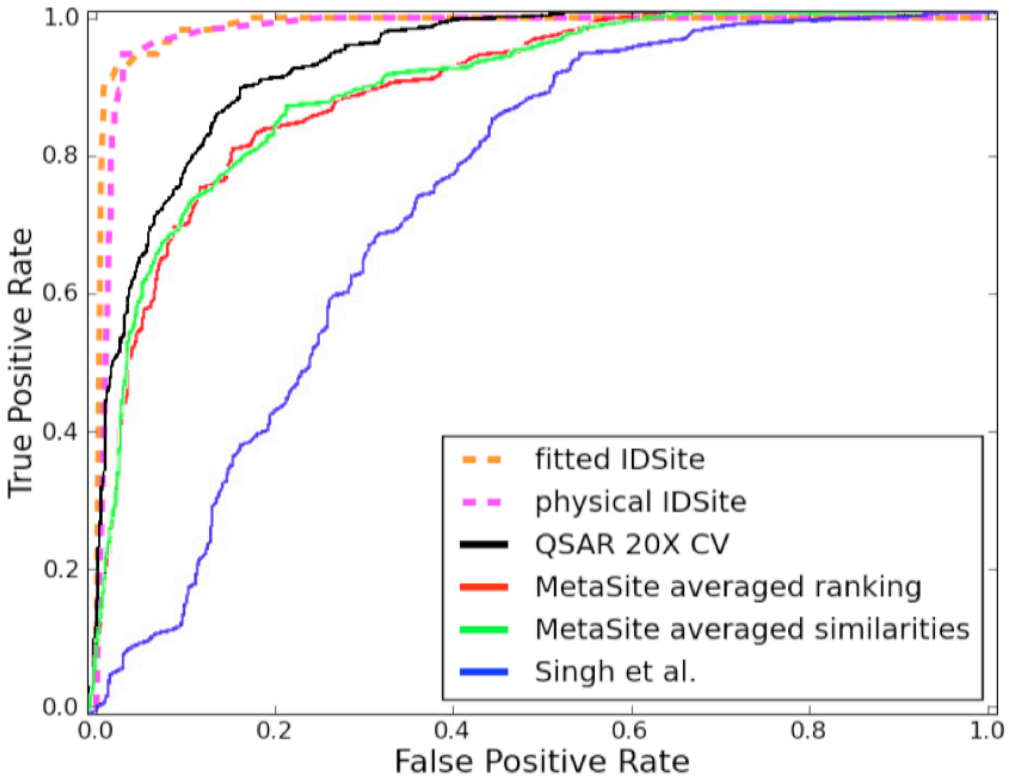


Figure 5.2. ROC curves for IDSite indicate its high predictive accuracy compared to other methodologies[141] used to predict P450 SOMs. It is important to note that the dataset used to produce the ROC curves for IDSite was different from the one used for the other methods.

6 Concluding Remarks

One major goal of my Ph. D. research was to promote the use of predictive computational chemistry for practical organic chemistry. Along the way towards this goal, Michelle Hall and I carried out the first fully quantum mechanical study of quantitative modeling of enantioselectivity with a large dataset. We initially investigated dioxirane catalyzed olefin epoxidation reactions in aqueous solvents and discovered the importance to optimize transition states within a solvation model for the accurate and robust prediction of experimental results. With structure optimizations in solvent and the B3LYP density functional, our model was able to predict the transition state free energy differences of the epoxidation reactions accurately, with a MUE < 0.5 kcal/mol. However, there still remained a few outliers in our dataset, which we attributed to the well-known deficiencies of B3LYP in describing dispersion interactions.

In order to improve the accuracy of our prediction protocol further, Dr. Arteum Bochevarov and I developed corrections for B3LYP, which significantly increase its accuracy for non-covalent interactions. Our newly developed correction scheme describes all types of non-covalent interactions accurately not only with large, but also with small basis sets. The accurate performance of our new methodology with small basis sets represents an important practical improvement over previously developed similar methodologies, which will be especially valuable when applied to the prediction of enantioselectivities.

More recently, I have now further improved the way in which the conformational searching of the transition states is carried out during the prediction of enantioselectivities. In addition, I also automated the setup of the predictions, such that

only a minimal amount of human work is now required to predict new reactions. Applications of our new prediction protocol to asymmetric CBS reductions are currently under way and will be reported in due time.

Moving from trying to predict the selectivity of organocatalyzed reactions, Dr. Jianing Li and I started trying to predicting reaction products of enzymatic reactions. Predicting enzymatic reactivity is a very challenging task, especially due to the high conformational flexibility of the enzyme catalysts. Nonetheless, we managed to obtain surprisingly accurate results for the regioselectivity prediction of metabolic reactions catalyzed by P450 2D6 enzymes with our newly developed IDSite methodology. Apart from making very accurate predictions based on a completely physical model, IDSite also produces accurate transition state like structures for the reactions in consideration, which could be useful starting guesses for QM/MM transition state optimizations. IDSite is currently being generalized to predict metabolites of other P450 enzyme families and shows promising preliminary results for a dataset of P450 2C9 substrates.

While the first part of this dissertation discussed some recent progress I made towards employing computers for making real world predictions of chemical reactions, the second part will focus on trying to gain further insight about the use of molecular wires in electrical devices, an approach which at some point in the future might provide access to more powerful computers than the ones currently available.

7 References

(1) Pedersen, S.; Herek, J. L.; Zewail, A. H. *Science* **1994**, *266*, 1359.

(2) Zewail, A. H. *Science* **1988**, *242*, 1645.

(3) Jaguar, *version 7.6*, Schrödinger, Inc., New York, NY, 2009.

(4) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 935.

(5) Eksterowicz, J. E.; Houk, K. N. *Chem. Rev.* **1993**, *93*, 2439.

(6) Titmuss, S. J.; Cummins, P. L.; Rendell, A. P.; Bliznyuk, A. A.; Gready, J. E. *J. Comput. Chem.* **2002**, *23*, 1314.

(7) Bredow, T.; Jug, K. *Theor. Chem. Acc.* **2004**, *113*, 1.

(8) Paldus, J.; Li, X. *Adv. Chem. Phys.* **1999**, *110*, 1.

(9) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **5612**, *98*, 5612.

(10) Koch, W.; Holthausen, M. C. *A chemist's guide to density functional theory*; 2 ed.; Wiley-VCH, 2001.

(11) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(12) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(13) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(14) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(15) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063.

(16) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374.

(17) Hobza, P.; Sponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315.

(18) Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624.

(19) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334.

(20) Tsuzuki, S.; Luthi, H. P. *J. Chem. Phys.* **2001**, *114*, 3949.

(21) Schneebeli, S. T.; Bochevarov, A. D.; Friesner, R. A. *J. Chem. Theory Comput.* **2011**, *7*, 658.

(22) Goldfeld, D. A.; Bochevarov, A. D.; Friesner, R. A. *J. Chem. Phys.* **2008**, *129*.

(23) Hall, M. L.; Goldfeld, D. A.; Bochevarov, A. D.; Friesner, R. A. *J. Chem. Theory Comput.* **2009**, *5*, 2996.

(24) Knoll, E. H.; Friesner, R. A. *J. Phys. Chem. B* **2006**, *110*, 18787.

(25) Rinaldo, D.; Tian, L.; Harvey, J. N.; Friesner, R. A. *J. Chem. Phys.* **2008**, *129*.

(26) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289.

(27) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

(28) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(29) Schneebeli, S. T.; Hall, M. L.; Breslow, R.; Friesner, R. *J. Am. Chem. Soc.* **2009**, *131*, 3965.

(30) Gladysz, J.; Michl, J. *Chem. Rev.* **1992**, *92*, 739.

(31) Enders, D.; Jaeger, K.-E. *Asymmetric Synthesis with Chemical and Biological Methods*; Wiley-VCH, 2007.

(32) Kagan, H. B.; Gopalaiah, K. *New J. Chem.* **2011**.

(33) Wong, O. A.; Shi, Y. *Chem. Rev.* **2008**, *108*, 3958.

(34) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.

(35) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*.

(36) Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P. *Chem. Rev.* **2010**, *110*, 5023.

(37) Corey, E. J.; Bakshi, R. K.; Shibata, S.; Chen, C. P.; Singh, V. K. *J. Am. Chem. Soc.* **1987**, *109*, 7925.

(38) Coleman, M. *Human Drug Metabolism: An Introduction*; John Wiley and Sons, 2010.

(39) Guengerich, F. P. *Annu. Rev. Pharmacol. Toxicol.* **2003**, *39*, 1.

(40) van de Waterbeemd, H.; Gifford, E. *Nat. Rev. Drug Discovery* **2003**, *2*, 192.

(41) Montellano, P. R. *Cytochrome P450: structure, mechanism, and biochemistry*; 4 ed.; Kluwer Academic/Plenum Publishers, New York, 2004.

(42) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.

(43) Oláh, J.; Mulholland, A. J.; Harvey, J. N. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6050.

(44) Li, J.; Abel, R.; Zhu, K.; Cao, Y.; Zhao, S.; Friesner, R. A. *Proteins: Struct., Funct., Bioinf.* **2011**, ASAP.

(45) Li, J. Towards high-resolution computational approaches for structure-based drug discovery. Ph.D. Thesis, Columbia University, New York, NY, 2011.

(46) Li, J.; Schneebeli, S. T.; Bylund, J.; Farid, R.; Friesner, R. A. *submitted*.

(47) Corbeil, C. R.; Thielges, S.; Schwartzentruber, J. A.; Moitessier, N. *Angew. Chem., Int. Ed.* **2008**, *47*, 2635.

(48) Moitessier, N.; Henry, C.; Len, C.; Chapleur, Y. *J. Org. Chem.* **2002**, *67*, 7275.

(49) MacroModel, *version 9.6*, Schrödinger, Inc., Portland, OR, 2007.

(50) Friesner, R. A. *Chem. Phys. Lett.* **1985**, *116*, 39.

(51) Friesner, R. A. *J. Chem. Phys.* **1986**, *85*, 1462.

(52) Friesner, R. A. *J. Chem. Phys.* **1987**, *86*, 3522.

(53) Friesner, R. A. *J. Phys. Chem.* **1988**, *92*, 3091.

(54) Martinez, T. J.; Carter, E. A. *Pseudospectral methods applied to the electron correlation problem in Modern Electronic Structure Theory*; World Scientific, Singapore, 1995; Vol. 2.

(55) Bach, R. D.; Coddens, B. A.; McDouall, J. J. W.; Schlegel, H. B.; Davis, F. A. *J. Org. Chem.* **1990**, *55*, 3325.

(56) Bach, R. D.; Dmitrenko, O.; Adam, W.; Schambony, S. *J. Am. Chem. Soc.* **2003**, *125*, 924.

(57) Bach, R. D.; Owensby, A. L.; Gonzalez, C.; Schlegel, H. B.; McDouall, J. J. W. *J. Am. Chem. Soc.* **1991**, *113*, 2338.

(58) Bach, R. D.; Owensby, A. L.; Gonzalez, C.; Schlegel, H. B.; McDouall, J. J. W. *J. Am. Chem. Soc.* **1991**, *113*, 6001.

(59) Bach, R. D.; Wolber, G. J. *J. Am. Chem. Soc.* **1984**, *106*, 1401.

(60) Dmitrenko, O.; Bach, R. D. *J. Phys. Chem. A* **2004**, *108*, 6886.

(61) Lang, T. J.; Wolber, G. J.; Bach, R. D. *J. Am. Chem. Soc.* **1981**, *103*, 3275.

(62) Armstrong, A.; Washington, I.; Houk, K. N. *J. Am. Chem. Soc.* **2000**, *122*, 6297.

(63) Houk, K. N.; Liu, J.; DeMello, N. C.; Condroski, K. R. *J. Am. Chem. Soc.* **1997**, *119*, 10147.

(64) Lucero, M. J.; Houk, K. N. *J. Org. Chem.* **1998**, *63*, 6973.

(65) Washington, I.; Houk, K. N. *Angew. Chem., Int. Ed.* **2001**, *40*, 4485.

(66) Christian, C. F.; Takeya, T.; Szymanski, M. J.; Singleton, D. A. *J. Org. Chem.* **2007**, *72*, 6183.

(67) Singleton, D. A.; Merrigan, S. R.; Liu, J.; Houk, K. N. *J. Am. Chem. Soc.* **1997**, *119*, 3385.

(68) Singleton, D. A.; Wang, Z. *J. Am. Chem. Soc.* **2005**, *127*, 6679.

(69) Singleton, D. A.; Wang, Z. H. *J. Am. Chem. Soc.* **2005**, *127*, 6679.

(70) Freccero, M.; Gandolfi, R.; Sarzi-Amadè, M.; Rastelli, A. *J. Org. Chem.* **2000**, *65*, 2030.

(71) Freccero, M.; Gandolfi, R.; Sarzi-Amadè, M.; Rastelli, A. *J. Org. Chem.* **2000**, *65*, 8948.

(72) Freccero, M.; Gandolfi, R.; Sarzi-Amadè, M.; Rastelli, A. *J. Org. Chem.* **2002**, *67*, 8519.

(73) Freccero, M.; Gandolfi, R.; Sarzi-Amadè, M.; Rastelli, A. *J. Org. Chem.* **2003**, *68*, 811.

(74) Freccero, M.; Gandolfi, R.; Sarzi-Amadè, M.; Rastelli, A. *J. Org. Chem.* **2004**, *69*, 7479.

(75) Freccero, M.; Gandolfi, R.; Sarzi-Amadè, M.; Rastelli, A. *J. Org. Chem.* **2005**, *70*, 9573.

(76) Freccero, M.; Gandolfi, R.; Sarziamade, M.; Rastelli, A. *Tetrahedron Lett.* **2001**, *42*, 2739.

(77) Markad, S. D.; Xia, S.; Snyder, N. L.; Surana, B.; Morton, M. D.; Hadad, C. M.; Peczuh, M. W. *J. Org. Chem.* **2008**, *73*, 6341.

(78) Wang, Z.-X.; Tu, Y.; Frohn, M.; Zhang, J.-R.; Shi, Y. *J. Am. Chem. Soc.* **1997**, *119*, 11224.

(79) Shi, Y. *Acc. Chem. Res.* **2004**, *37*, 488.

(80) Tian, H.; She, X.; Shu, L.; Yu, H.; Shi, Y. *J. Am. Chem. Soc.* **2000**, *122*, 11551.

(81) Tian, H.; She, X.; Xu, J.; Shi, Y. *Org. Lett.* **2001**, *3*, 1929.

(82) MacroModel, *version 6.0*, Schrödinger, Inc., Portland, OR, 1998.

(83) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(84) Venables, D. S.; Schmuttenmaer, C. A. *J. Chem. Phys.* **1998**, *108*, 4935.

(85) Ahnercan, G.; Krienke, H.; Schmeer, G. *J. Mol. Liq.* **2006**, *129*, 75.

(86) MacroModel, *version 9.6*, Schrödinger, Inc., Portland, OR, 2007.

(87) Armstrong, A.; Ahmed, G.; Dominguez-Fernandez, B.; Hayter, B. R.; Wailes, J. S. *J. Org. Chem.* **2002**, *67*, 8610.

(88) Solladie-Cavallo, A.; Jerry, L.; Klein, A.; Schmitt, M.; Welter, R. *Tetrahedron: Asymmetry* **2004**, *15*, 3891.

(89) Denmark, S. E.; Matsuhashi, H. *J. Org. Chem.* **2002**, *67*, 3479.

(90) Armstrong, A.; Dominguez-Fernandez, B.; Tsuchiya, T. *Tetrahedron* **2006**, *62*, 6614.

(91) Armstrong, A.; Moss, W. O.; Reeves, J. R. *Tetrahedron: Asymmetry* **2001**, *12*, 2779.

(92) Armstrong, A.; Tsuchiya, T. *Tetrahedron* **2006**, *62*, 257.

(93) Wang, Z. X.; Miller, S. M.; Anderson, O. P.; Shi, Y. *J. Org. Chem.* **2001**, *66*, 521.

(94) Burke, C. P.; Shi, Y. *Angew. Chem., Int. Ed.* **2006**, *45*, 4475.

(95) Silva, M. A.; Bellenie, B. R.; Goodman, J. M. *Org. Lett.* **2004**, *6*, 2559.

(96) Bach, R. D.; Glukhovtsev, M. N.; Gonzalez, C. *J. Am. Chem. Soc.* **1998**, *120*, 9902.

(97) Cheong, P. H. Y.; Yun, H.; Danishefsky, S. J.; Houk, K. N. *Org. Lett.* **2006**, *8*, 1513.

(98) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.

(99) Crehuet, R.; Anglada, J. M.; Cremer, D.; Bofill, J. M. *J. Phys. Chem. A* **2002**, *106*, 3917.

(100) Duefert, A.; Werz, D. B. *J. Org. Chem.* **2008**, *73*, 5514.

(101) R. C. Parr and W. Yang *Density Functional Theory of Atoms and Molecules,* Oxford University Press, New York, 1989.

(102) Johnson, E. R.; Mackie, I. D.; DiLabio, G. A. *J. Phys. Org. Chem.* **2009**, *22*, 1127.

(103) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2722.

(104) Grimme, S.; Antony, J.; Schwabe, T.; Muck-Lichtenfeld, C. *Org. Biomol. Chem.* **2007**, *5*, 741.

(105) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press, Oxford, U. K., 1997.

(106) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23.

(107) Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291.

(108) Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Chem. Soc., Perkin Trans. 2* **2001**, 651.

(109) Grimme, S.; Muck-Lichtenfeld, C.; Antony, J. *J. Phys. Chem. C* **2007**, *111*, 11199.

(110) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2007**, *127*.

(111) Chai, J. D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

(112) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397.

(113) Grafenstein, J.; Cremer, D. *J. Chem. Phys.* **2009**, *130*.

(114) Sato, T.; Nakai, H. *J. Chem. Phys.* **2009**, *131*.

(115) Rezac, J.; Jurecka, P.; Riley, K. E.; Cerny, J.; Valdes, H.; Pluhackova, K.; Berka, K.; Rezac, T.; Pitonak, M.; Vondrasek, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2008**, *73*, 1261.

(116) Valdes, H.; Pluhackova, K.; Pitonak, M.; Rezac, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747.

(117) Liu, B.; McLean, A. D. *J. Chem. Phys.* **1973**, *59*, 4557.

(118) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(119) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.

(120) Pittner, J.; Hobza, P. *Chem. Phys. Lett.* **2004**, *390*, 496.

(121) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690.

(122) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *J. Chem. Phys.* **2005**, *122*.

(123) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2000**, *122*, 3746.

(124) Headgordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503.

(125) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(126) Pople, J. A.; Headgordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(127) Scuseria, G. E.; Schaefer, H. F. *J. Chem. Phys.* **1989**, *90*, 3700.

(128) Pitonak, M.; Janowski, T.; Neogrady, P.; Pulay, P.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1761.

(129) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*.

(130) Jaguar, *version 7.6*, Schrödinger, Inc., New York, NY, 2009.

(131) Kaminski J. R.; Murphy, R. B. B., D. A.; Friesner, R. A., G A.; Maple *J. Chem. Theory Comput.* **2005**, *1*, 248.

(132) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(133) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(134) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98*, 1358.

(135) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1994**, *100*, 2975.

(136) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.

(137) Allemann, C.; Um, J. M.; Houk, K. N. *J. Mol. Catal. A: Chem.* **2010**, *324*, 31.

(138) Calderon, F.; Doyagueez, E. G.; Cheong, P. H.-Y.; Fernandez-Mayoralas, A.; Houk, K. N. *J. Org. Chem.* **2008**, *73*, 7916.

(139) Zhu, H.; Clemente, F. R.; Houk, K. N.; Meyer, M. P. *J. Am. Chem. Soc.* **2009**, *131*, 1632.

(140) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. *J. Med. Chem.* **1999**, *42*, 1515.

(141) Sheridan, R. P.; Korzekwa, K. R.; Torres, R. A.; Walker, M. J. *J. Med. Chem.* **2007**, *50*, 3173.

Part II. Chemistry for Computers: From novel antiaromatic and π-stacked molecular wires to highly conducting link groups with direct Au-C bonds

1 Background and Introduction

1.1 Electronic devices made of single molecules

The exponential increase in the speed and memory capacity of computer chips observed during the past several decades also commonly known as Moore's Law[1] has been possible thanks to the ability of the computer industry to produce smaller and smaller transistors. With smaller transistors in hand the computer factories were able to put an increasing number of transistors into CPU and memory chips, which lead to the observed exponential increase in computational capacity at an almost constant hardware cost. Since the chip sizes will eventually hit the size limits of single molecules, the trends predicted by Moore's law are likely to slow down at some point in the future, unless new chemical techniques can be developed, which allow the use of electrical wires at the molecular scale for computer hardware.

Moore's Law has been a major driving force of our world's economy for the last couple of decades.[2] Furthermore, it is also at least partially responsible for the increasing success of predictive computational chemistry, which has significantly advanced our understanding of complicated chemical processes and has helped chemists finding better chemical solutions faster. Therefore, it seems of great importance to keep the trends of Moore's law going or to even better speed them up further.

In order to move forward along the lines of creating devices at the single molecule level, it is fundamental to gain a deeper understanding of the relationship between the electronic properties of single molecule devices and the molecular structures from which they are constructed.[3,4] Professor M. Ratner has been among the first to propose the use

of single molecules in electrical devices back in 1974.[5] While experimental realizations of such single molecule devices were not possible back then, this has changed with the more recent development of atomic level microscope techniques.[6]

During the past decade, a number of novel ways have been developed to link molecules to electrodes to produce single molecule electrical wires.[7] One of the most successful current approaches is to use molecules with link groups, which bind to gold electrodes formed with either the STM break junction technique or the mechanically controlled break junction technique.[8-10] Figure 1.1 illustrates how a molecular circuit is formed with the STM break junction technique. In a first step the STM gold tip is smashed into a gold surface. Then, the tip is slowly pulled away from the surface again. During that process, a gold wire is being pulled out of the gold surface, which becomes thinner and thinner and finally breaks to leave a gap at the molecular scale. Sample molecules, which are simply swimming around the gold electrodes in a solvent (usually 1,2,4-trichlorobenzene), are then able to bridge the gap while binding to both electrodes at the same time with gold binding groups, such as amines.

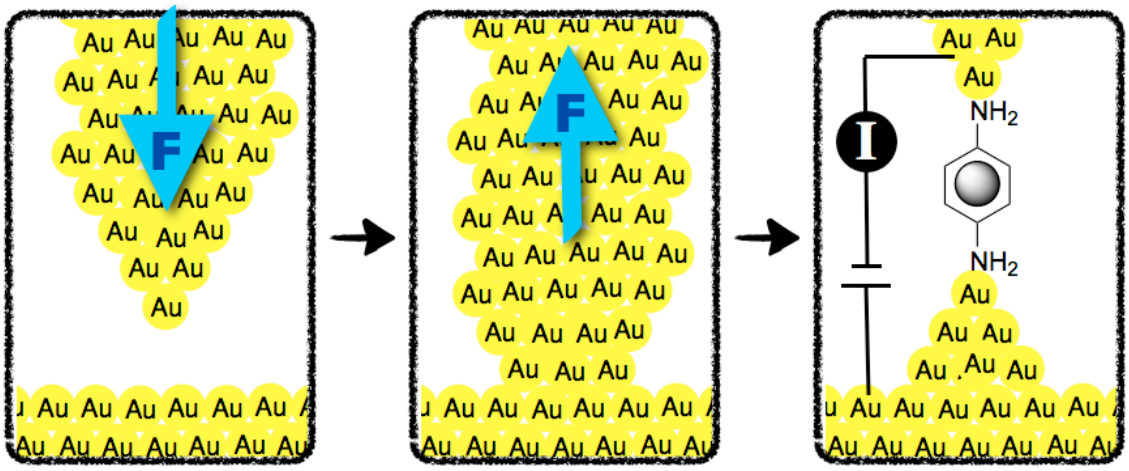


Figure 1.1. How a STM gold break junction is formed. Step 1: The STM gold tip is smashed into the surface. Step 2: The STM tip is pulled out from the surface to produce a thin gold wire connecting the two electrodes. Step 3: The gold wire eventually breaks and leaves a gap at the molecular scale in which a molecule with gold binding groups (e.g. amines) can bind.

One major advantage of the STM based gold break junction technique is that the single molecule conductance measurements can easily be repeated thousands of times, such that the single molecule conductance of molecules can be measured with statistical significance. Statistical averaging is more difficult to achieve with other techniques, such as using carbon nanotubes cut with an oxygen plasma or an electron beam as the electrodes, since usually only a hand-full of functional devices can be produced in reasonable time.[11-13]

Many different molecular systems have been proposed to function as highly conducting molecular wires bridging gaps between two metal electrodes, including oligo(phenylene-ethinylene), oligo(phenylene-vinylene), polyphenyls and polythiophenes.[14-19] These molecules form extended conjugated systems, where the delocalization of the electrons through the molecular backbone decreases the energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied

molecular orbital (LUMO). Electron transport measurements through such molecules bonded to metal electrodes with various different chemical link groups including thiols and amines have been performed, primarily with the STM based break-junction technique.[9,20-23]

1.2 Aromatic and Antiaromatic effects on conductance

Earlier work on diaminoacenes in the Breslow and Venkataraman labs had shown that 9,10-diaminoanthracene was more conducting than 1,4-diaminonaphthalene, which was a better conductor than 1,4-diaminobenzene (Figure 1.2).[24]

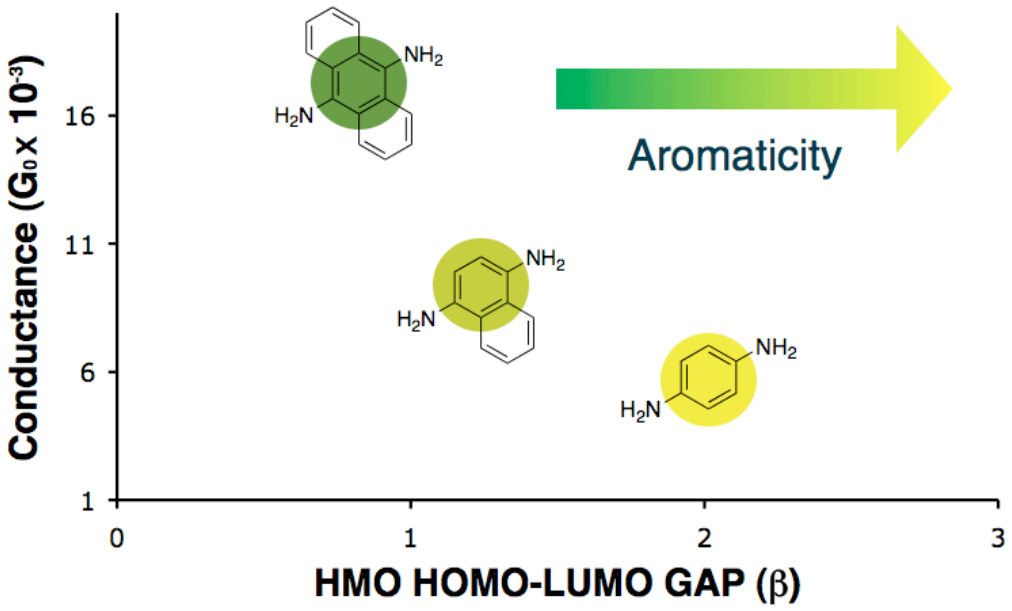


Figure 1.2. Aromaticity fights conductance. Shown is a plot of single molecule conductance determined with gold STM break junction measurements against the Hückel Molecular Orbital (HMO) HOMO-LUMO gaps of benzene, naphtalene and anthracene. β is a parameter of the HMO theory, describing the Hamiltonian element between two neighboring π-orbitals.

In this series, the trends in the HOMO-LUMO gap correlated with the conductance. In addition, since the N in these molecules could form partial pi-bonds with the adjacent C when bonded in a junction, developing some quinoid-character in the ring, it was interesting to note that the conductance of these diamino-acenes also correlated with the stability of the corresponding quinones and semiquinones relative to their hydroquinones.[25] The observed correlation of semiquinone stabilities with conductance can be rationalized by considering that oxidation of the electron rich diaminoacenes is much easier than their reduction. Therefore, diamine linked molecular wires conduct mainly through their HOMOs rather than through their LUMOs.[24] Conduction through the HOMO means that during electron transport the molecule in the break junction gets partially oxidized before a new electron gets injected from the negative electrode (anode). Therefore, the molecule in the junction develops partial semiquinone character during electron transport through the HOMO, which makes molecules with lower oxidation potential better conductors.[25] Since the stability of the semiquinones is determined mainly by the aromaticity of the system (Figure 1.3), with the least aromatic compound, anthracene, forming the most stable semiquinone, increased aromaticity leads to reduced conductance.

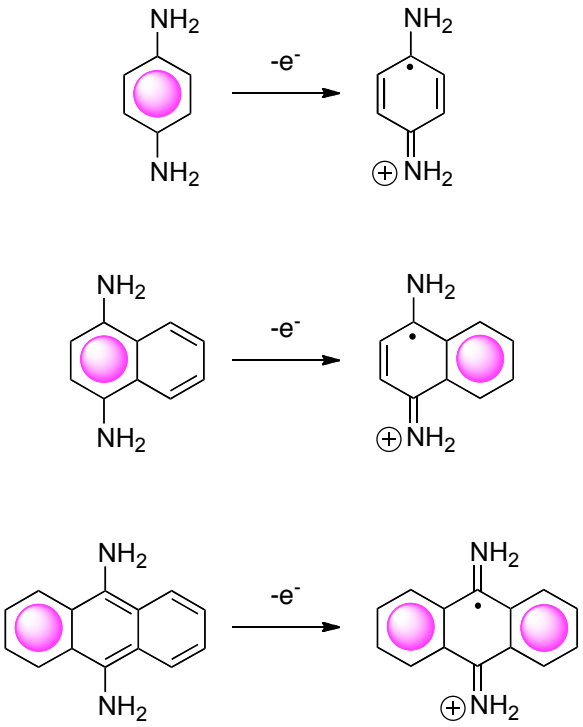


Figure 1.3. The effects of aromaticity on the semiquinone stabilities of benzene, naphtalene and anthracene diamine. Benzenediamine is the hardest to oxidize as the aromaticity gets lost when the semiquinone is formed. In contrast, 9,10-Diaminoanthracene easily loses an electron as one benzenoid ring is "gained" during the process.

This result suggests the possibility of increased conductance in antiaromatic compounds,[26] where the equivalent of a semiquinoid form with diminished antiaromaticity could be favorable.[27-29] This idea is illustrated for the partially antiaromatic 2,7-diaminobiphenylene in Figure 1.4.

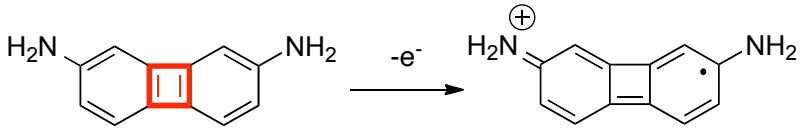


Figure 1.4. Oxidation of 2,7-diaminobiphenylene to the semiquinone removes the antiaromaticity present in the system and should therefore lead to increased conductance. 2,7-diaminobiphenylene is shown in the least stable resonance form, which shows the antiaromatic cyclobutadiene scaffold in red.

Furthermore, the HOMO-LUMO gap is generally smaller in antiaromatic rings than in aromatic ones, and thus one would expect a higher conductance for antiaromatic compounds.

1.3 Conductance through pi-pi stacked systems

One area of particular interest for electron transport studies is how π-π stacked aromatic rings conduct electricity, since electron transport through stacked DNA base pairs[30] and amino acid residues[31] plays a key role in xenobiotic metabolism,[32] photosynthesis[33] and DNA repair.[34] Inspired by how π-π stacking is efficiently used by nature to achieve directed long range electron transport, it is believed that synthetic molecular wires incorporating π-π stacking will be useful as new electronic materials[35] or as components in molecular devices.[36,37] Furthermore, rigid synthetic molecular wires represent model systems for their biological counterparts, for which it is often difficult to obtain reproducible conductance data due to the large size and the increased flexibility.[38] Measurements of molecular layers incorporating a paracyclophane scaffold[39,40] and conductance between molecules held together by spontaneous π-π stacking[16,41] have provided evidence for transport across two π-systems. However the important question of how the electron transport properties of multiple π-stacked systems differ from saturated or conjugated chains had not been addressed experimentally at the single molecule level yet. Therefore, we decided to further study the conductance of such systems at the single molecule level and were indeed able to do so by creating direct Au-C contact between strained multiple layered cyclophanes and gold electrodes.[42] Further details of this study will be revealed in chapter 3 of this dissertation.

1.4 The need for alternative linker groups with high conductance

The first linker groups used to attach single molecules to gold electrodes were all quite strong Lewis bases such as amines, phosphines, sulfides and selenides.[21,22] While such linker groups lead to well defined conductance traces of the single molecules bound in the junctions, there is generally a relatively large resistance associated with the electrical links formed between the gold electrodes and the molecules. The resistance stemming from the gold-molecule links is also referred to as the Shottky-barrier of the molecular junction.[43]

The large Shottky-barriers observed with the aforementioned traditional lewis basic link groups, significantly increase the resistance of the resulting molecular junctions. Therefore it is desirable to look for alternative linker groups, which would display lower Shottky-barriers. The last two chapters of this dissertation will therefore describe two alternative approaches to link molecules to gold electrodes, which were just recently developed in the Breslow and Venkataraman research groups. Both approaches represent a way to directly link gold electrodes to carbon atoms of the molecular wires and therefore eliminate the need for Lewis basic heteroatom binding groups and their associated Shottky-barriers.

2 Single molecule conductance through antiaromatic molecular wires

2.1 Previous attempts to study antiaromatic effects on electrical conductance

Antiaromatic cyclobutadienes would provide ideal test molecules to study antiaromatic effects on molecular wires, but simple cyclobutadienes perform Diels-Alder reactions on themselves so rapidly that they cannot be handled under normal room temperature conditions. Thus initial studies focused on the more stable biphenylene system **1**, with the cyclobutadiene ring fused with benzene rings. The proposed antiaromatic effect on conductivity could be considerably smaller in biphenylene, reflecting the greatly diminished cyclobutadiene character in the central ring. In fact only one resonance form out of four has both double bonds in the four-membered ring (Figure 2.1), and that one is a minor contributor.

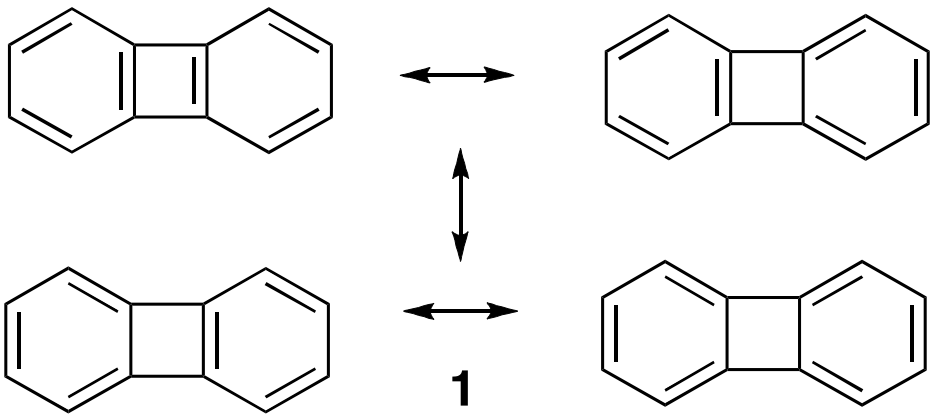


Figure 2.1. Resonance forms of biphenylene.

Since amino groups proved to be reliable gold binding groups for STM break junction conductance measurements, which couple well into the pi-systems of aromatic

molecules and bind selectively to undercoordinated gold,[20,25] previous work in the Breslow and Venkataraman groups was focused on trying to measure the single molecule conductance of 2,7-diaminobiphenylene (**2**). As a "non-aromatic" comparison system for the partially antiaromatic biphenylene, 2,7-diaminofluorene (**3**) had been chosen. In contrast to biphenylene, which formally contains an antiaromatic cyclobutadiene in one of its resonance forms, fluorene contains a "non-aromatic" cyclopentadiene core and therefore represents a reasonable comparison system to estimate the effects of the antiaromaticity present in biphenylene on its conductance (see also Figure 2.2).

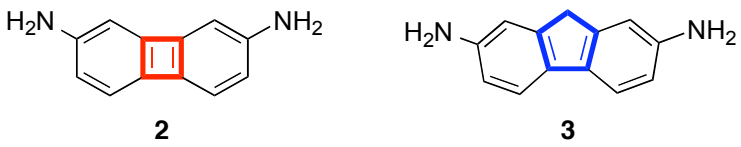


Figure 2.2. The partially antiaromatic biphenylene molecular wire **2** and its "non-aromatic" fluorene comparison compound **3**. The antiaromatic cyclobutadiene core of **2** is highlighted in red, while the "non-aromatic" cyclopentadiene core of **3** is shown in blue.

2,7-Diaminobiphenylene **2** was already a known compound, and was synthesized following the published procedure,[44] while 2,7-diaminofluorene (**3**) was simply purchased from Sigma-Aldrich. In reference 44, the authors reported a melting point and a low resolution mass spectrum for 2,7-diaminobiphenylene (**2**), but no NMR spectra. Dr. Frank Foss, a former postdoctoral fellow in the Breslow group, was able to obtain a ^{1}H NMR spectrum of an impure sample of **2**, which proved to be unstable. While the authors of reference 44 reported that solid **2** turned color on standing in air, Dr. Foss found that oxidation occurs instantly in air with solutions of **2**. He and Professor Venkataraman attempted cyclic voltammetry and conductivity measurements on

reasonably pure solutions of **2** in an inert atmosphere, but the CV curves were not reversible and the conductivity was ill defined at best. Even though the STM conductance measurements were carried out under inert gas atmosphere, **2** presumably still got oxidized at the positive gold electrode of the measurement device, which then led to uncontrolled decomposition. As biphenylene (**1**) itself is a perfectly air-stable compound, even at elevated temperatures, the instability of compound **2** must be caused by the two strongly electron donating amine gold binding groups. Thus we turned to a biphenylene derivative and its fluorene analog with contacts that were less electron-donating, which finally resulted in an air stable biphenylene derivative, which gave well defined single molecule conductance traces and also showed two reversible oxidation peaks in the CV. The synthesis and electrical properties of this air-stable biphenylene molecular wire and its electrical properties are described in the following sections.

2.2 An airstable partially antiaromatic molecular wire

As described in the previous section, amino groups were too electron donating to be used as linkers to gold with biphenylene molecular wires, rendered the resulting 2,7-diaminobiphenylene (**2**) unstable. As an alternative to amine linker groups, the Nuckolls and Venkataraman groups had successfully employed thioether groups as linkers to gold electrodes as well.[22] However, while thiomethyl groups showed well-defined conductance peaks when attached to aliphatic chains, they gave broad conductance peaks when attached to aromatic systems. The broadening of the conductance histograms observed for thiomethyl groups coupled into conjugated systems is due to the relatively low torsional barriers of the aromatic C-S bonds, and can be avoided by replacing the

thiomethyl groups with "locked" cyclic thioether derivatives.[45] The main reason for the reduced torsional barrier of an aromatic C-S bond compared to the one of an aromatic C-N bond is the larger size of the sulfur lone pair orbitals, which therefore couple less strongly into the aromatic system and lead to longer C-S bonds. As a direct consequence, thioether groups are also significantly less electron donating than amino groups. Therefore, thioether gold binding groups seemed ideal candidates for antiaromatic, air-stable biphenylene molecular wires.

As the synthesis of "locked" cyclic thioether derivatives usually involves significant synthetic effort, we initially thought of simpler alternative thioether compounds that might allow us to determine antiaromatic effects on conductance. In collaboration with Dr. Rachid Skouta, a former postdoc from the Breslow group and Maria Kamenetska, a graduate student in the Venkataraman group, we discovered that molecules with benzylic thioether link-groups showed well defined conductance peaks, although the conductance values observed with these link groups were relatively low. The low conductance caused by the presence of the extra methylene groups is mainly due to the disruption of the direct π-conjugation between the Au-S bonds and the conjugated aromatic systems. Since we were only interested in relative conductance values, the low conductivities obtained with benzylic thiomethyl gold binding groups did not pose a problem for our intent of using such binding groups to study antiaromatic effects on conductance. However, we were a little worried that the large Shottky-barriers observed with the benzylic thiomethyl link groups would completely dominate the overall conductance of our molecular wires and therefore mask any conductance changes brought upon the systems by changing the aromatic part of the conductance pathways.

Therefore, before using benzylic thiomethyl binding groups to study antiaromatic effects on conductance, we first studied a few model systems to confirm that the expected trends in conductance are clearly seen even in the presence of the additional benzylic methylene groups.

The model systems we looked are were the para substituted benzene derivative **4**, as well as the corresponding 1,4 and 9,10 substituted naphtalene and anthracene derivatives **5** and **6**. Compounds **4** – **6** were either purchased from commercial sources or synthesized according to known procedures.[46,47] In full agreement with the trends observed in the corresponding diamine series,[25] we find that the anthracene compound **6** conducts higher than the naphtalene compound **5**, which is itself a better conductor than the corresponding benzene derivative **4**. A comparison of the conductance trends observed in the diamine and the benzylic thiomethyl series is shown in Figure 2.3.

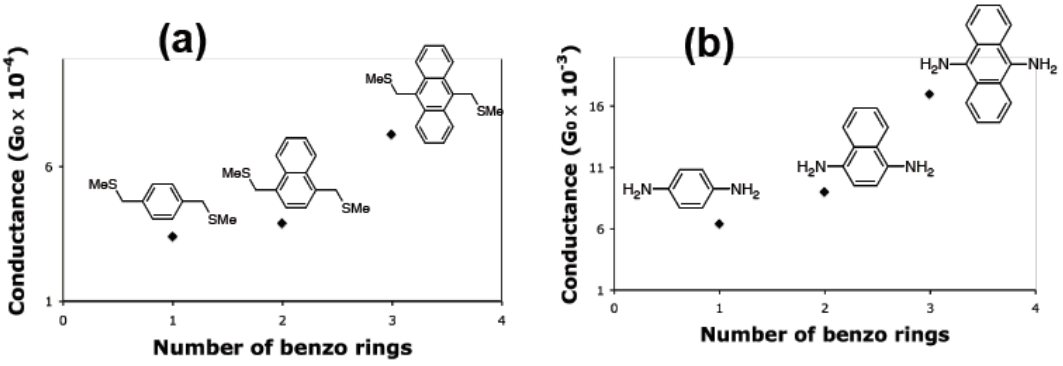


Figure 2.3. Almost identical trends for single molecule conductance are observed through the benzene, naphtalene, anthracene series with (a) benzylic thiomethyl link groups and (b) amine[25] link groups.

Single molecule conductance traces for compounds **4** – **6** are shown in Figure 2.4. Apart from a well-defined major peak, all conductance traces of compounds studied with

benzylic thiomethyl gold binding groups show an additional shoulder peak at higher conductance, the origin of which is still under investigation. One plausible explanation for the origin of these higher conducting shoulder peaks would be that one of the two gold electrodes contacts the π-system of the aromatic rings directly while the other side of the molecule is anchored to the second gold electrode with the sulfur atom.

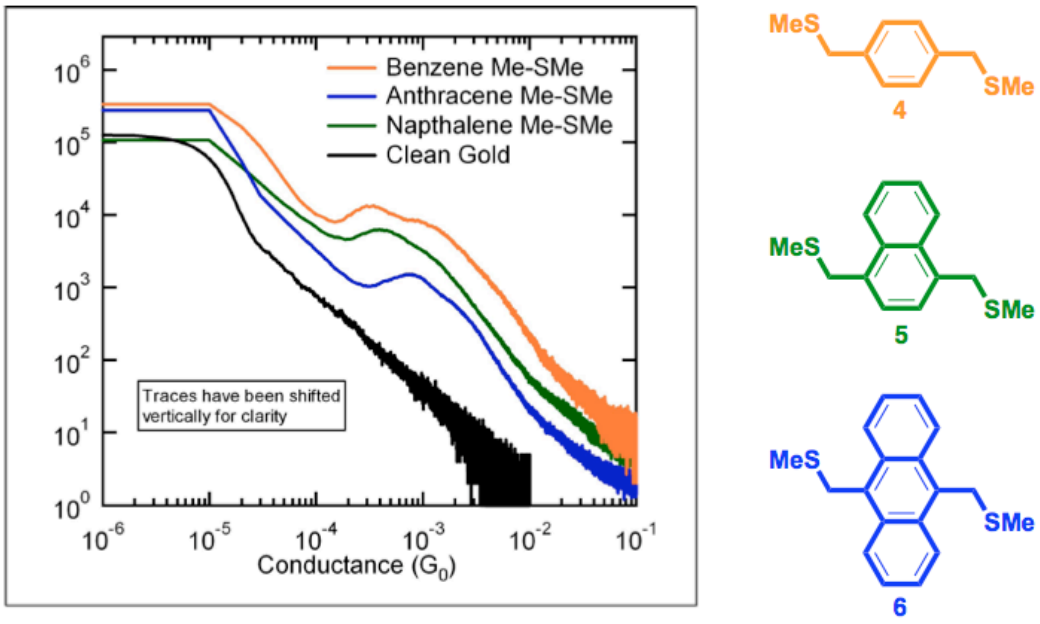


Figure 2.4. Single molecule conductance histograms of benzene, naphtalene and anthracene measured with benzylic thiomethyl binding groups. The conductance values for the major peaks are $3.4 \cdot 10^{-4}$ G_0 for **4**, $3.9 \cdot 10^{-4}$ G_0 for **5** and $7.2 \cdot 10^{-4}$ G_0 for **6**. Note the shoulder peak at higher conductance than the main peak, which is present in all conductance traces. The origin of this peak is still under investigation.

As the trends in conductance values observed with benzylic thiomethyl binding groups almost perfectly mirrored the trends obtained with amino groups for the benzene, naphtalene and anthracene series, we decided to hang the benzylic thiomethyl binding groups onto our biphenylene molecular wire (compound **7**) to study the effects of antiaromaticity on single molecule conductance. The two main advantages of the benzylic thiomethyl binding groups over other alternatives are the high stability of the

resulting molecular wires as well as the straightforward syntheses of the compounds, which are described in the following section.

To our surprise, the biphenylene molecular wire **7** with benzylic thiomethyl binding groups showed a very broad conductance peak in the single molecule conductance histogram (Figure 2.5), which made it impossible to fit an accurate conductance value for that molecule. The origin of the broad conductance peak is unclear, it is however possible that there are multiple binding configurations of the biphenylene compound with the gold electrodes, some of which might involve direct contacts between the gold electrode(s) and the π-system of the biphenylene wire. Forming direct Au-C contacts with the biphenylene molecular wire seem more likely than in the case of the corresponding fluorenyl wire, since the π-electrons in the biphenylene wire are destabilized due to the partial antiaromaticity in the biphenylene system. Therefore, the antiaromatic destabilization of the π-electrons in the biphenylene wire could be responsible for the broad conductance peak observed in that case.

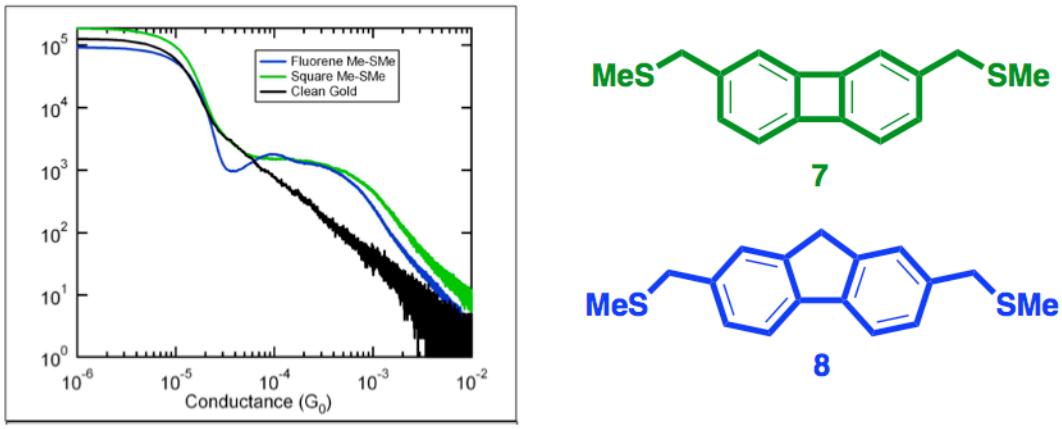


Figure 2.5. Single molecule conductance histograms of the partially antiaromatic biphenylene molecular wire **7** and its corresponding fluorenyl analogue **8** with a "non-aromatic" cyclopentadiene core. While the fluorene compound **8** showed a conductance of $1.0 \bullet 10^{-4}$ G_0, the conductance peak for the biphenylene **7** was very broad, such that no main conductance value could be obtained. (Conductance measurements were done by Maria Kamenetska in the Venkataraman group, while Dr. Rachid Skouta from the Breslow group synthesized the fluorene compound **8**).

Since no accurate conductance value could be obtained for the biphenylene molecular wire **7**, we decided to replace the benzylic thiomethyl gold binding groups with cyclic thioethers with sulfurs directly attached to the aromatic system. Such cyclic thioether derivatives were found to show very well defined conductance peaks with relatively high conductance by the Nuckolls and Venkataraman groups.[45] However, while Professor Nuckolls and coworkers used five membered cyclic thioethers, we decided to use six memembered rings instead, simply because six membered rings seemed much easier to synthesize. We therefore designed the biphenylene molecular wire **9a** and its corresponding fluorene analogue **10a** to finalize the study of antiaromaticity on single molecule conductance.

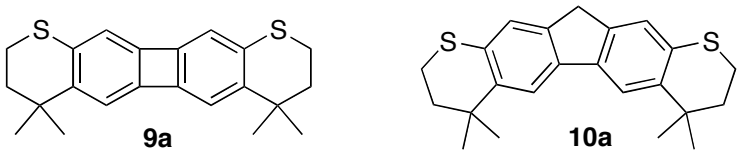
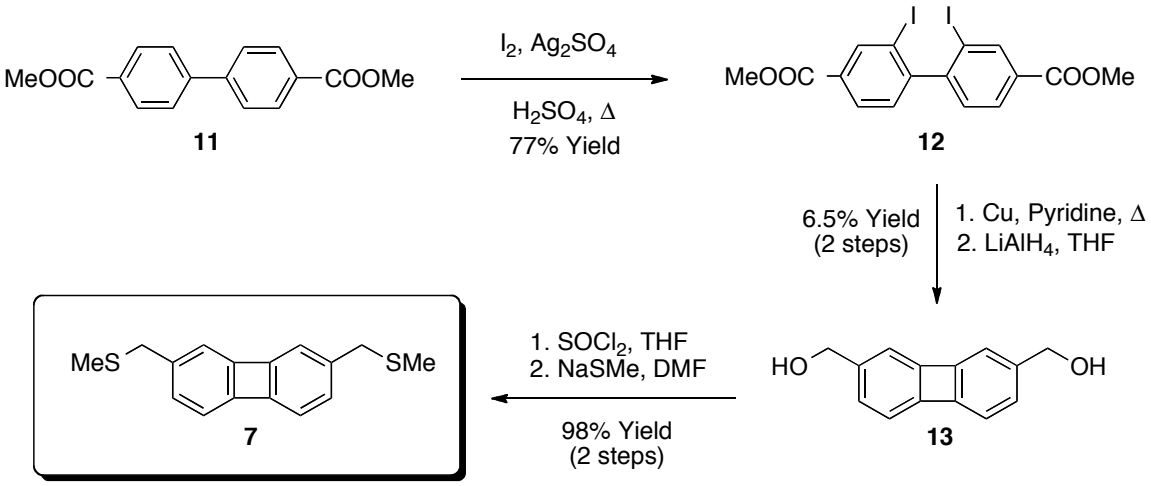


Figure 2.6. Biphenylene (**9**) and fluorene (**10a**) molecular wires with cyclic thioether link groups.

Our syntheses of the molecular wires **9a** and **10a** as well as their single molecule and bulk electrical properties are described in the following sections.

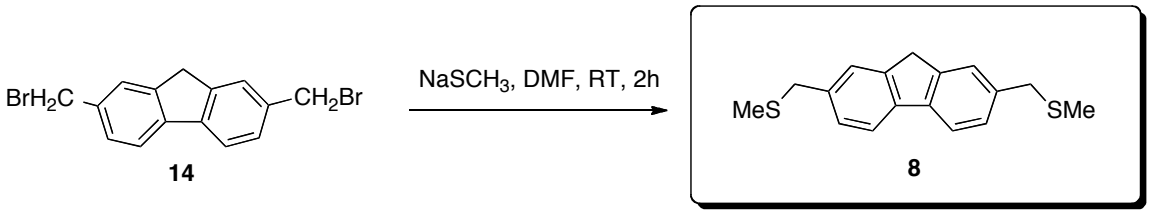
2.3 Synthesis of the biphenylene molecular wires

Scheme 2.1 shows the synthesis of the biphenylene molecular wire **7** with benzylic thiomethyl gold binding groups. Our synthesis started with commercially available **11**, which was converted to the biphenylene derivative **13** in low yield following a known procedure.[48] The diol **13** was then converted to the dichloride and reacted in situ with sodium thiomethoxide to give the final product **7** in acceptable yield.



Scheme 2.1. Synthesis of the air-stable biphenylene molecular wire **7**.

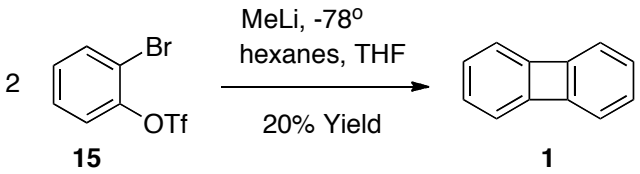
The synthesis of the fluorene derivative **8** was carried out by Dr. Rachid Skouta. As shown in Scheme 2.2, he simply reacted the known 2,7-bis(bromomethyl)fluorene[49] (**14**) with sodium thiomethoxide and obtained the desired fluorene derivative **8** in almost quantitative yield.



14 **8**

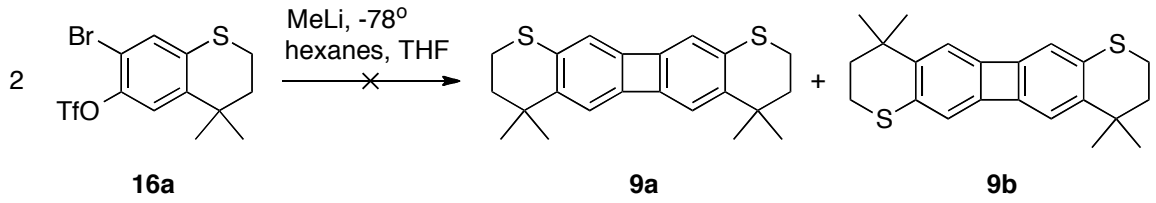
Scheme 2.2. Synthesis of the fluorene molecular wire **8** with benzylic thiomethyl binding groups by Dr. Rachid Skouta.

While the synthesis of the biphenylene molecular wire **7** with benzylic thiomethyl gold binding groups was relatively easy, a lot more effort went into the synthesis of the more sophisticated biphenylene derivative **9a** with "locked" cyclic thioether electrode contacts. Our initial attempt to prepare the biphenylene derivative **9a** involved a benzyne dimerization as the key step. While we were able to obtain unsubstituted biphenylene in modest yield via a benzyne dimerization reaction starting from the ortho-bromotriflate **15** (a reaction found by Dr. Zhanling Cheng, a former postdoc in the Breslow lab) (Scheme 2.3), we never detected even a trace of the biphenylene molecular wire **9a** or its regioisomer **9b** following similar procedures (Scheme 2.4).



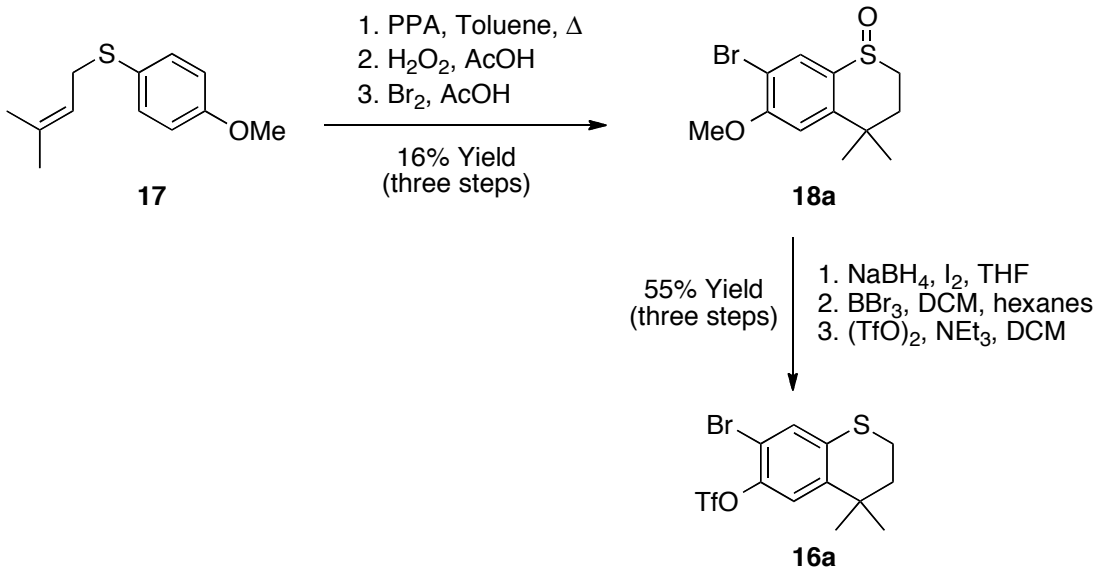
15 **1**

Scheme 2.3. Synthesis of biphenylene **1** via a benzyne dimerization mechanism (Dr. Zhanling Cheng).



Scheme 2.4. Attempted synthesis of the biphenylene molecular wire **9a** and its regioisomer 9b via a benzyne dimerization.

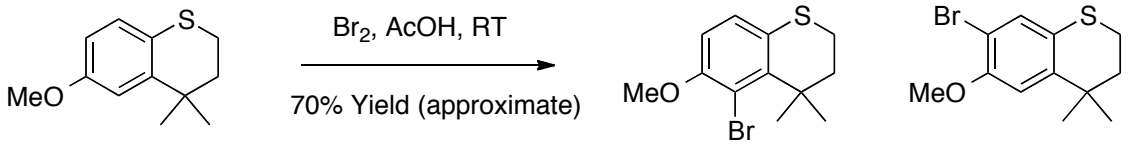
The precursor for the attempted benzyne dimerization shown in Scheme 2.4 was prepared in low overall yield as shown in Scheme 2.5.



Scheme 2.5. Synthesis of the o-bromotriflate precursor (**16a**) for the benzyne dimerization.

Protection of the cyclic sulfide **19** as the sulfoxide proved necessary as direct bromination of the free sulfide with bromine in acetic acid gave the undesired regioisomer **20b** as the major bromination product (Scheme 2.6). This surprising regioselectivity is most likely caused by a mechanism, where the sulfide initially

becomes brominated to the bromonium ion. We believe that the bromonium ion itself then acts as the brominating agent and transfers the electrophilic bromine atom to the benzene ring of another molecule via the bimolecular transition state shown in Figure 2.7. Simple electrostatics should favor the assembly shown in Figure 2.7, where the bromonium ion lies closely above the methoxy group. Such a transition state assembly would then lead to the observed major regioisomer (**20b**) of the reaction. One fact, which supports such a mechanism is that the cyclic sulfide starting material disappears almost immediately after addition of bromine to the reaction mixture (a TLC of the reaction mixture taken 5 minutes after addition of bromine revealed no more sulfide starting material), but formation of the final reaction product takes much longer (up to 48 hours). Also, if the reaction is worked up with water shortly after addition of the bromine, the major product detected by mass spectrometry is the sulfoxide of **19**, which presumably was formed by hydrolysis of the bromonium ion intermediate.



19 **20b** **major** **20a** **minor**

Scheme 2.6. Direct bromination of the cyclic sulfide **19** surprisingly gave the undesired regioisomer as the major product. From ^{1}H NMR the ratio of **20b**:**20a** was estimated as 2.4:1.0.

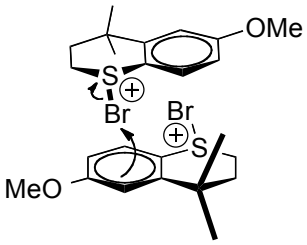
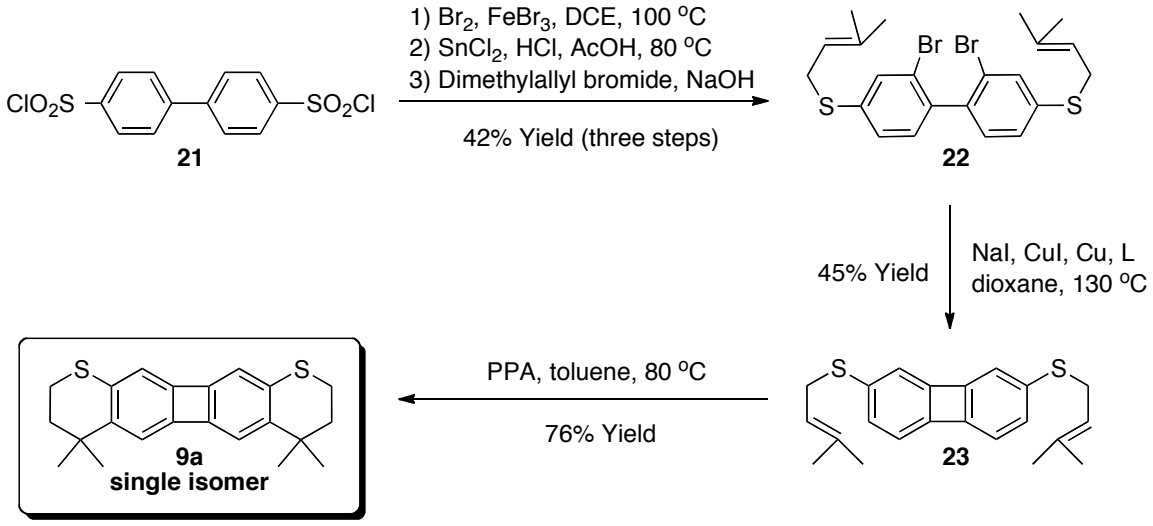


Figure 2.7. Transition state assembly of bromonium ion intermediates, which could explain the surprising regioselectivity (and the slow reaction rate) observed in the direct bromination reaction of compound **19**.

If the above mentioned bimolecular reaction mechanism is really responsible for the formation of the undesired regioisomer **20b**, one should be able to obtain the desired isomer by adding an excess of bromine to the reaction mixture. Reaction conditions with excess bromine were tried, but did not give satisfactory results mainly due to multiple bromination and uncontrolled decomposition. Therefore, protection of the sulfide as the sulfoxide before bromination (as shown in Scheme 2.5) seemed the best solution to the problem.

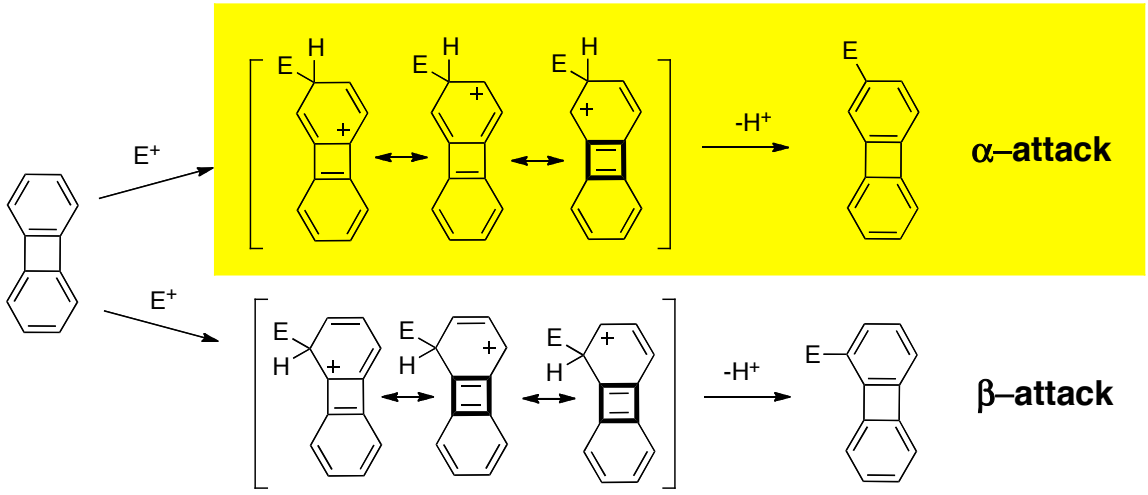
Since we were not able to get the benzyne dimerization of compound **16a** to work, we designed an alternative synthesis of our biphenylene molecular wire **9a**. The alternative synthesis shown in Scheme 2.7 started from commercially available biphenyldisulfonyl chloride **21**. **21** was brominated twice under harsh conditions (excess bromine with $FeCl_3$ in a sealed tube at 100°C for 6 days) and then reduced it *in situ* to the bis thiol to which we attached two dimethylallyl units. The resulting bis-bromobiphenyl **22** was then converted into the diiodide following a methodology developed by Buchwald and Klapars,[50] and under the reaction conditions the biphenylene compound **23** was formed directly in an Ullman-type coupling. Finally, we then cyclized **23** under acidic conditions to **9a**. In principle **23** might cyclize to either carbon 1 or carbon 3, and

at the other end to carbon 8 or carbon 6, but the cyclization reaction was completely regioselective, cyclizing only onto carbons 3 and 6.



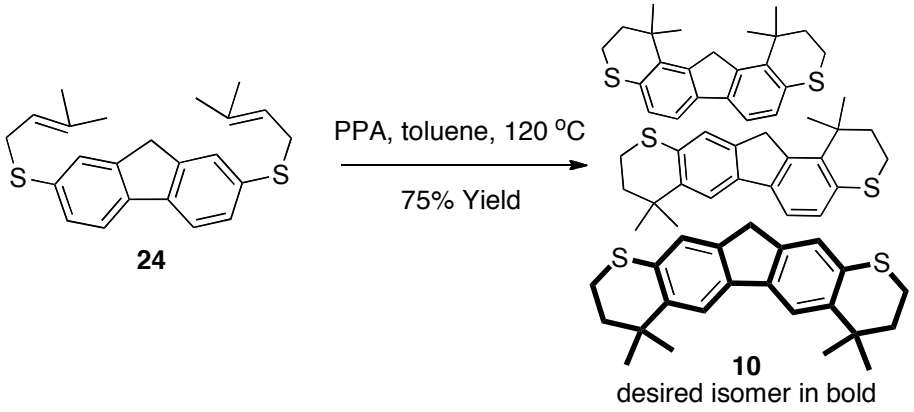
Scheme 2.7. Highly selective synthesis of the biphenylene molecular wire **9a**.[51] L is *trans*-N,N'-dimethylcyclohexanediamine.

This selectivity reflects the result of antiaromaticity in the four-membered ring.[52,53] Biphenylenes preferentially substitute on the β-carbon,[54] in contrast with naphthalenes that substitute on the alpha carbon.[55] In naphthalene the attachment to carbon 1 is used in order to produce reasonable stabilization of a cation without disrupting the aromaticity of the second benzene ring. The preference to go to the β-carbon in the biphenylene system reflects the fact that disrupting the antiaromaticity of the four-membered ring is desirable and occurs in two of the resonance forms for β-attack (Scheme 2.8). Only one favorable resonance form can be drawn for α-attack, making β-substitution preferred.



Scheme 2.8. α-electrophilic attack at biphenylenes (highlighted in yellow) is favored because of the antiaromatic character of the molecule. Major resonance forms of the intermediate for α and β electrophilic attack on biphenylene are shown (with the bottom ring fixed). Destabilizing cyclobutadiene parts are shown in bold.

The synthesis of the fluorene derivative **10a** shown in Scheme 2.9 followed a similar route as the synthesis of biphenylene **9a** and was carried out by Dr. Rachid Skouta.



Scheme 2.9. Synthesis of a mixture of fluorenyl molecular wires **10** by Dr. Rachid Skouta.[51]

As expected, the cyclization reaction of **24** gave a mixture of isomers, which I separated with preparative HPLC.

Biphenylene **9a** was much easier to handle than the diamino analog **2** and did not oxidize in air. This allowed us to determine its electrical properties at the single molecule level and in the bulk. The results of these electrical studies on our biphenylene molecular wire **9a** are described in the following section.

2.4 The electrical properties of a biphenylene

In the past the Breslow and Venkataraman groups had found that derivatives of benzene with lower oxidation potential had higher conductivities, suggesting that in the conducting state the ring becomes partially p-doped as some electron density is withdrawn by the anode before electrons could be fed in from the cathode.[25,26] This trend had been verified by measuring oxidation potentials of diamine molecular wires using Cyclic Voltammetry (CV). Therefore, we first recorded the CV traces of our biphenylene and fluorene molecular wires **9a** and **10a** in acetonitrile at room temperatue. The CV traces (Figure 2.8) showed two reversible oxidations for each molecule and indicate that the biphenylene **9a** is indeed more easily oxidized (by ca. 120 mV) than its fluorenyl analogue **10a** as expected from the antiaromatic character present in the biphenylene compound.

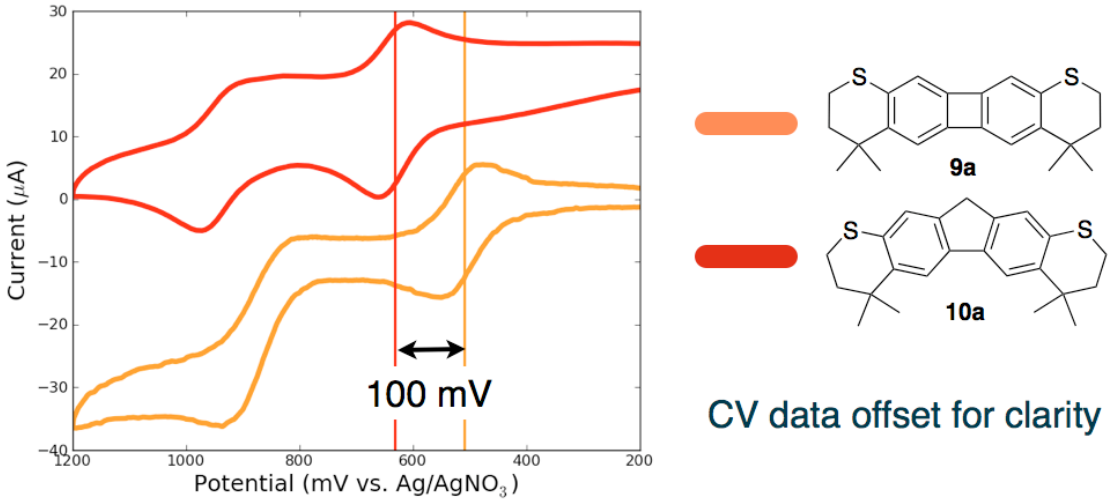


Figure 2.8. CV traces of **9a** and **10a** in acetonitrile at Room Temperature. Both compounds show two reversible oxidation waves. The biphenylene molecular wire **9a** is more easily oxidized (by ca. 100 mV) due to its antiaromatic character.[51] The orange trace was lowered for clarity from its original starting point at zero current.

As in the diamine series compounds with lower oxidation potentials were generally better conductors, we were expecting that our biphenylene molecular wire **9a** might conduct higher than its corresponding fluorenyl analogue **10a**.

Maria Kamenetska from the Venkataraman lab therefore measured single molecule conductance traces of our biphenylene and fluorene molecular wires using the STM based gold break-junction technique. Single-molecule junctions were created by repeatedly forming and breaking gold point contacts in a solution of the molecules in 1,2,4-trichlorobenzene in a home-built set-up.[56] For each molecule studied, the measured conductance traces reveal steps at molecule-dependent conductance values less than the quantum of conductance, $G_0 = 2e^2/h$; these are due to conduction through a molecule bonded in the gap between the two Au point contacts. Figure 2.9 shows conductance histograms generated (without any data selection) from over 5000 measured traces with each molecule. As mentioned earlier, measurements of **2** were carried out in an argon

atmosphere due to its instability in air, but even so it gave a broad conductance range that could not be identified with a single value. In contrast, biphenylene **9a** with locked cyclic thioether gold binding groups showed a well defined conductance value of $(3.6 \pm 0.2) \cdot 10^{-3}\ G_0$, which was however not significantly different from the one of the corresponding fluorene reference compound ($(3.5 \pm 0.3) \cdot 10^{-3}\ G_0$) with a non-aromatic core.[51]

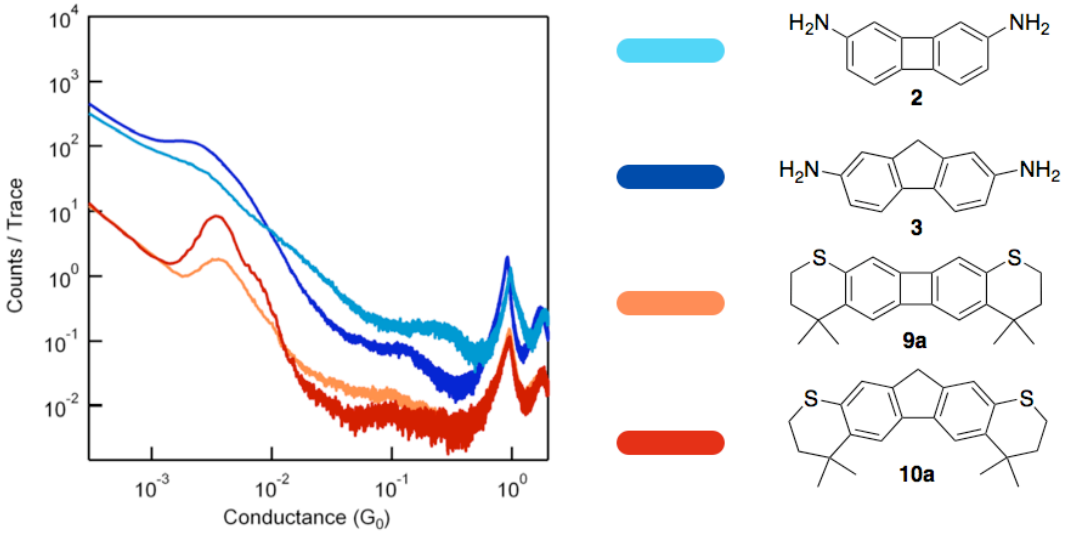


Figure 2.9. Linearly binned conductance histograms for partially antiaromatic biphenylene molecular wires and their corresponding fluorenyl analogues with "non-aromatic" cyclopentadiene cores.[51]

To gain a better understanding of these results, in particular the reason that oxidation potentials did not correlate with conductivities, Dr. Mark Hybertsen and Dr. Hector Vazquez performed DFT calculations for the two diamines (**2, 3**) and the two cyclic thioether compounds (**9a**, **10a**). In each case the geometry was fully optimized. In agreement with overall expectations, the calculated HOMO-LUMO gaps for the

biphenylene species are about 20% smaller than those for the flourene species. However, the calculated vertical ionization potentials are very similar, with the biphenylene species showing values 0.1 eV smaller, at most. The tunnel coupling was estimated using a single gold atom to model the attachment point of each linker to its respective electrode. The tunnel coupling is proportional to the splitting between the symmetric and antisymmetric combinations of the Au-molecule-Au frontier orbitals. This general method to assess tunnel coupling and conductance[57,58] has been shown to correctly predict the trends in low-bias conductance of single gold-diamine-molecule junctions.[20] The predicted ratio of the conductance through the amine-linked biphenylene **2** to that through the amine linked flourene **3** was 1.3 and the predicted ratio for the methylsulfide linked **9a** to that for **10a** was 1.2. These modest changes in conductance follow the trend suggested by the HOMO-LUMO gap trends and are only slightly larger than the observed experimental conductance differences.

2.5 Effects of the structure of biphenylene on its conductance

As discussed in the previous section, we found that our biphenylene molecular wire showed the same conductance within experimental error as our non-aromatic reference molecular wire with a fluorenyl backbone.[51] Since the reduced oxidation potential of the biphenylene compared to the fluorene wire should have increased its conductance, there must be an additional effect causing a lowering of the biphenylene conductance, such that the two molecules finally end up about the same. In fact, the lowering of the biphenylene conductance could simply be caused by its special structure, which also reflects the antiaromaticity of the system. While cyclobutadiene itself exists exclusively in an elongated rectangular form, in which the two double bonds can be as far

apart as possible, this is also true in the biphenylene system. An X-ray crystal structure of biphenylene[59] (Figure 2.10), indicates that the single bonds between the two benzene rings in biphenylene have a length of 1.514 Å, quite close to the bond distance for an ordinary single bond between sp^3 carbons, rather than the normal shorter bond between sp^2 carbons. Such a bond length should make the conductivity worse, since passing electrons from one benzene ring to the other requires crossing what is a rather long distance for π-conjugation.[60] As another striking result, the benzene rings in biphenylene (Figure 2.10) are almost classical cyclohexatriene rings with alternating long and short bonds, in order to decrease the π electron density in the four-membered ring. In a sense the antiaromaticity in the cyclobutadiene ring has partially dearomatized two benzene rings. Such situations have been remarked on previously.[61]

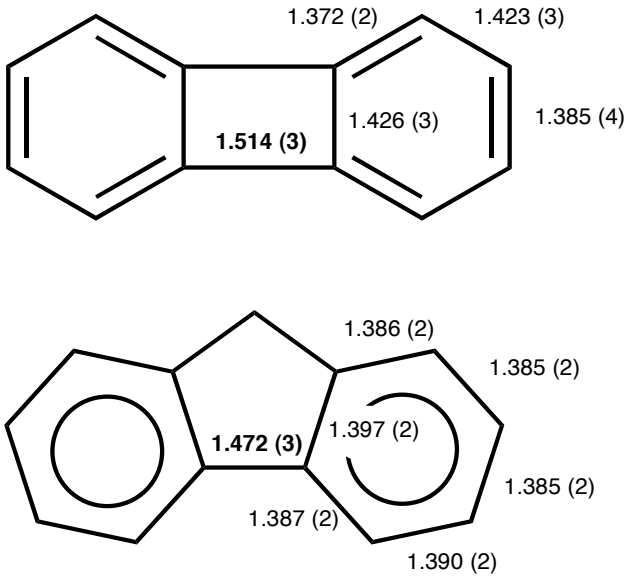


Figure 2.10. The structures of biphenylene[59] and fluorene[62] determined by X-ray crystallography indicate the very long bonds between the two benzenes in biphenylene and the cyclohexatriene-like character of the benzene rings, with alternating bond lengths.

Thus we think the best summary of what our study of biphenylenes has shown is that a significant degree of antiaromaticity is indeed present in the central ring of biphenylene, as demonstrated by the unusual aromatic substitution pattern, the distorted structure, and the easy oxidation of the system, but it does not show up in conductivity because there are two competing considerations. One is that the semiquinoid character of the system when it conducts electricity will diminish the cyclobutadiene character of the central ring; that was the effect we were expecting to lead to better conductivity. The other is that the conductivity should be worse because of the long bonds connecting the two benzene rings, and perhaps the bond length alternation in the benzene rings. The two effects apparently more or less compensate, but they would also be partly present in a simple cyclobutadiene derivative. Thus there seems to be little hope for high conductivity in cyclobutadiene-based molecular wires.

2.6 Experimental details

General procedures. *Chemicals:* All solvents, inorganic salts, and organic reagents were purchased from commercial sources and used without further purification unless otherwise mentioned. Compounds **5**[46], **8**[51], **10**[51], **14**[51] and **24**[51] were prepared by Dr. Rachid Skouta, while Dr. Frank Foss prepared compound **2**[44] and Dr. Zhanling Cheng prepared compound **15**[63]. Compounds **6**[47], **12**[64], **13**[48] and **17**[65] were prepared following known procedures by me.

Chromatography: Merck pre-coated 0.25 mm silica plates containing a 254 nm fluorescence indicator were used for analytical thin-layer chromatography. Flash chromatography was performed on 230-400 mesh silica (SiliaFlash® P60) from Silicycle. Preparative HPLC was run on a Waters 600 liquid chromatography system

equipped with a Waters™ 600 pumping system and a Waters 2489 UV-Vis detector. Samples were collected manually. A Waters XBridge™ C18 reverse phase preparative column (particle size 5 μm, 19x150 mm) was used as the stationary phase.

Spectroscopy: NMR spectra were obtained on a Bruker DPX 300 or 400 MHz spectrometer. Spectra were analyzed with the MestreNova Software (Version 6.1). Chemical shifts for the proton spectra are reported in ppm (δ) relative to chloroform (δ 7.26). Chemical shifts for the carbon spectra are reported in ppm (δ) relative to chloroform (δ 77.16) as well. CI-MS spectra were taken on a Nermag R-10-10 instrument. FAB MS spectra were taken on a JEOL JMS-DX-303 HF instrument using either glycerol or p-nitrobenzyl alcohol as matrices. Matrix assisted laser desorption ionization (MALDI) mass spectra were acquired using an Applied Biosystems Voyager DE Pro time-of-flight mass spectrometer. Positive ion mass spectra were acquired in the linear mode using a nitrogen laser (337nm). Instrument settings were as follows: accelerating voltage, 21,000 volts; grid voltage, 95%; guide wire, 0.05%; extraction delay time, 200 nsec. All data processing was performed using Applied Biosystems Data Explorer v 4.0.0.0.

Synthesis of 2,7-Bis(methylthiomethyl)biphenylene (7). 2,7-Bis(hydroxymethyl)biphenylene (**13**[48], 22 mg, 0.10 mmol), which was carefully dried on the high vacuum before use, was dissolved in anhydrous THF (3.0 ml) under Argon atmosphere. The clear solution was then cooled to 0º C and thionylchloride (0.1 ml, 0.42 mmol) was added dropwise. The reaction mixture was then allowed to warm to room temperature and stirred for 20 hours. The solvent as well as the excess of thionylchloride was then distilled off under reduced pressure and a solution of sodium thiomethoxide (30 mg,

0.31 mmol) in about 2.5 ml anhydrous DMF was added to the crude biphenylene dichloro compound under Argon atmosphere. After the addition of the sodium thiomethoxide, the color of the reaction mixture immediately turned to red and the reaction mixture was stirred for 20 hours at room temperature. Then, the reaction mixture was diluted with ethyl acetate, washed with H_2O, dried over $MgSO_4$ and the solvent was evaporated under reduced pressure. The crude product was purified by flash colum chromatography on silica gel (10% DCM in hexanes) to afford biphenylene **7** in quantitative yield as a yellow solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.68 (s, 2H), 6.62 (d, J = 7.0 Hz, 2H), 6.56 (d, J = 7.0 Hz, 2H), 3.47 (s, 4H), 2.02 (s, 6H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 151.33, 149.56, 138.51, 128.54, 118.55, 116.93, 39.08, 14.79; LRMS (FAB+): 272 $[M]^+$; HRMS calcd for $C_{16}H_{16}S_2$: 272.0693; found 272.0705.

Synthesis of 7-bromo-6-methoxy-4,4-dimethylthiochroman 1-oxide (18a). Compound **17**[65] (8.32 g, 40.0 mmol) was dissolved in Toluene (100 ml) and PPA (15 ml, warmed to about 60º C in order to make it less viscous) was added at room temperature. Then, the biphasic reaction mixture was refluxed for 5 hours at 120º C under argon atmosphere with vigorous stirring. The stirring was then continued at 70º C for additional 5 days. Finally, the reaction mixture was cooled to room temperature, the toluene was decanted and the PPA phase was extracted three times with Et_2O. The combined organic fractions were then evaporated under reduced pressure to afford crude 6-methoxy-4,4-dimethylthiochroman (**19**) in almost quantitative yield as a yellow oil. The crude product was reasonably pure (as judged by ^{1}H NMR) and was therefore used without further purification. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.00 (d, J = 8.6 Hz, 1H), 6.94 (d, J = 2.8 Hz,

1H), 6.66 (dd, *J* = 8.6, 2.8 Hz, 1H), 3.77 (s, 3H), 3.22 – 2.82 (m, 2H), 2.11 – 1.77 (m, 2H), 1.32 (s, 6H).

The crude 6-methoxy-4,4-dimethylthiochroman (**19**, 7.73 g, 37.1 mmol) was then dissolved in ca. 20 ml acetic acid and H_2O_2 (35 wt% soln. in H_2O, 3.0 ml, 45 mmol) was slowly added at room temperature. After 5 minutes of vigorous stirring, the reaction mixture was diluted with ca. 150 ml H_2O and extracted three times with DCM. The combined organic layers were then washed with saturated aqueous $NaHCO_3$ solution until neutral pH. Then, the organic layers were dried over $MgSO_4$, filtered, evaporated under reduced pressure and dried *in vacuo* to give crude 6-methoxy-4,4-dimethylthiochroman 1-oxide in almost quantitative yield, which was used for the following bromination reaction without further purification. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.65 (d, *J* = 8.6 Hz, 1H), 6.97 (d, *J* = 2.6 Hz, 1H), 6.87 (dd, *J* = 8.6, 2.6 Hz, 1H), 3.84 (s, 3H), 3.29 – 2.91 (m, 2H), 2.62 (ddd, *J* = 14.6, 11.1, 3.4 Hz, 1H), 1.81 (ddd, *J* = 14.9, 6.5, 2.8 Hz, 1H), 1.46 (s, 3H), 1.28 (s, 3H).

Crude 6-methoxy-4,4-dimethylthiochroman 1-oxide (0.91 g, 4.06 mmol) was dissolved in ca. 10 ml acetic acid. The solution was then cooled to 0º C and Br_2 (0.21 ml, 4.1 mmol) was slowly added at that temperature. Then, the frozen reaction mixture was slowly warmed to room temperature and stirred for 2 hours. The reaction was then quenched with 5% aqueous $Na_2S_2O_3$ solution, diluted with H_2O and extracted with DCM. The combined organic layers were then washed with saturated aqueous $NaHCO_3$ solution until neutral pH, dried over $MgSO_4$, filtered and evaporated under reduced pressure. As the crude product was unstable, it was immediately purified with flash column chromatography (15 – 20% EA in DCM) over silica gel to give 7-bromo-6-methoxy-4,4-

dimethylthiochroman 1-oxide (**18a**) in low yield (16% over three steps). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.86 (s, 1H), 6.90 (s, 1H), 3.93 (s, 3H), 3.21 – 2.97 (m, 2H), 2.67 – 2.42 (m, 1H), 1.92 – 1.75 (m, 1H), 1.46 (s, 3H), 1.30 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 158.57, 146.05, 135.60, 116.90, 114.08, 110.73, 56.51, 42.92, 34.74, 31.39, 31.06, 28.87; LRMS (APCI+): 303 and 305 $[M+1]^+$. Compound **18** was obtained as an inseparable 2:1 mixture of the regioisomers 7-bromo-6-methoxy-4,4-dimethylthiochroman 1-oxide and 5-bromo-6-methoxy-4,4-dimethylthiochroman 1-oxide. The mixture was carried forward as such and separated at the stage of the orthobromotriflate **16**.

Synthesis of the benzyne-dimerization precursor 16a. To a solution of 7-bromo-6-methoxy-4,4-dimethylthiochroman 1-oxide (**18a**, 0.188 g, 0.620 mmol) in anhydrous THF (5 ml) under Argon atmosphere was added I_2 (0.144 g, 0.567 mmol) and $NaBH_4$ (0.038 g, 1.00 mmol). The reaction mixture was stirred for ca. 5 min at room temperature and then quenched carefully with 1 N aqueous NaOH solution and 5% aqueous $Na_2S_2O_3$ solution. Then, the reaction mixture was diluted with about 20 ml H_2O and extracted with DCM. The combined organic phases were then dried over $MgSO_4$, filtered and evaporated under reduced pressure to give crude 7-bromo-6-methoxy-4,4-dimethylthiochroman in good yield. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.28 (s, 1H), 6.92 (s, 1H), 3.87 (s, 3H), 3.12 – 2.93 (m, 2H), 2.06 – 1.83 (m, 2H), 1.34 (s, 6H); LRMS (FAB+): 286 and 288 $[M]^+$; HRMS calcd for $C_{12}H_{15}O^{79}BrS$: 286.0027; found 286.0028. Since the starting material was already a 2:1 mixture of regioisomers, this compound was also obtained as an inseparable 2:1 mixture of 7-bromo-6-methoxy-4,4-dimethylthiochroman

and 5-bromo-6-methoxy-4,4-dimethylthiochroman. The mixture was carried foreward as such and separated at the stage of the triflate **16**.

The hydrolysis of the methoxy group of 7-bromo-6-methoxy-4,4-dimethylthiochroman was carried out in the following way: A solution of 7-bromo-6-methoxy-4,4-dimethylthiochroman and 5-bromo-6-methoxy-4,4-dimethylthiochroman (2:1 mixture, 0.21 g, 0.73 mmol) in anhydrous DCM under Argon atmosphere was cooled to -78º C. Then, BBr_3 (1.0 M solution in hexanes, 5.0 ml, 5.0 mmol) was added to the reaction mixture dropwise at -78º C. After the addition was complete, the reaction mixture was warmed to room temperature and was stirred for an additional 22 hours at that temperature. Then, excess BBr_3 was carefully quenched with H_2O at -78º C and the reaction mixture was diluted with H_2O. After extraction with DCM, the combined organic phases were dried over MgSO4, filtered and concentrated under reduced pressure to give crude 7-bromo-4,4-dimethylthiochroman-6-ol. ^1H NMR (300 MHz, $CDCl_3$) δ 7.17 (s, 1H), 7.05 (s, 1H), 5.21 (s broad, 1H), 3.10 – 2.91 (m, 2H), 1.96 – 1.88 (m, 2H), 1.30 (s, 6H).

Triflate protection of the phenol was then carried out according to the following procedure: To a solution of 7-bromo-4,4-dimethylthiochroman-6-ol (0.190 g, 0.695 mmol) in 7 ml anhydrous DCM under Argon atmosphere was added NEt_3 (0.23 ml, 1.7 mmol) and trifluoromethanesulfonic anhydride (0.14 ml, 0.83 mmol) at 0º C. The reaction mixture was then stirred for two hours at 0º C and for an additional two hours at room temperature. As a TLC still showed some left over starting material, additional NEt_3 (0.2 ml, 1.7 mmol) and trifluoromethanesulfonic anhydride (0.1 ml, 0.8 mmol) were added and the reaction was stirred at room temperature for an additional 30 minutes.

Then, the reaction mixture was concentrated under reduced pressure to a black oil, which was purified with flash column chromatography (10% DCM in hexanes) over silica gel to give the orthobromotriflate **16a** in 55% overall yield over three steps. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.37 (s, 1H), 7.27 (s, 1H), 3.10 – 2.92 (m, 2H), 2.16 – 1.83 (m, 2H), 1.32 (s, 6H). LRMS (FAB+): 404 and 406 $[M]^+$; HRMS calcd for $C_{12}H_{12}O_3{}^{79}BrF_3S_2$: 403.9363; found 403.9376.

Synthesis of (2,2'-dibromobiphenyl-4,4'-diyl)bis((3-methylbut-2-enyl)sulfane) (22). Commercially available 4,4'-biphenyldisulfonyl chloride (**21,** 5.00 g, 14.2 mmol) was dissolved in dry DCE (25 ml). Bromine (8 ml, 155 mmol) and iron(III) bromide (0.3 g, 1.0 mmol) were added and the reaction mixture was stirred at 100 °C for 8 days in a high pressure sealed tube with a teflon screw-cap. The reaction mixture was then allowed to cool to room temperature. 1.0 ml of the reaction mixture (out of about 33 ml) were transferred to a 25 ml round bottom flask and excess bromine as well as the DCE were evaporated on the rotavap. The dark brown oily residue was dried on the high vacuum for 2 hours. The residue (about 0.43 mmol, assuming 100% yield in the first stage and a 1/33 dilution after the first stage) was dissolved in a HCl solution (1.0 M, 5.0 ml) in acetic acid and anhydrous stannous chloride (1.75 g, 9.23 mmol) was added. Then, the reaction mixture was then stirred at 80 °C for 1 hour. During the reaction time, the reaction mixture became colorless. The reaction mixture was then cooled to room temperature and the solvent was evaporated under reduced pressure. The residue was dissolved in 10 ml EtOH and a solution of NaOH (0.50 g, 12.5 mmol) in 1 ml 50 % EtOH/H_2O was slowly added at room temperature under argon. The reaction mixture was

then stirred for 5 min at room temperature, 3,3-dimethylallyl bromide (0.13 ml, 1.11 mmol) was added and the reaction mixture was warmed to 45 °C for 1 hour under argon. The solvent was evaporated on the rotavap and the residue was taken up in DCM. The organic phase was washed with H_2O, dried over $MgSO_4$, filtered and concentrated under reduced pressure. The resulting oil was purified by flash colum chromatography on silica gel (10% DCM in hexanes) to afford **22** in 45% overall yield over 3 steps (assuming a 1/33 dilution after the first stage) as a colorless oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.59 (d, *J* = 1.8 Hz, 2H), 7.29 (dd, *J* = 8.1, 1.9 Hz, 4H), 7.10 (d, *J* = 8.0 Hz, 2H), 5.52 – 5.22 (m, 2H), 3.59 (d, *J* = 7.5 Hz, 4H), 1.75 (s, 6H), 1.64 (s, 6H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.13, 138.99, 137.39, 132.50, 131.12, 127.88, 123.83, 118.81, 32.11, 25.83, 17.99; LRMS (FAB+): 510 $[M]^+$; HRMS calcd for $C_{22}H_{24}Br_2S_2$: 509.9686; found 509.9709.

Synthesis of 2,7-bis(3-methylbut-2-enylthio)biphenylene (23). A mixture of **22** (100 mg, 0.195 mmol), CuI (35 mg, 0.184 mmol), NaI (100 mg, 0.667 mmol) and copper powder (100 mg, 1.574 mmol) was dried for 12 hours under high vacuum in a 10 ml high pressure reaction vessel. The reaction vessel was then backfilled three times with argon and the reagents were suspended in anhydrous dioxane (2 ml). *trans*-N,N'-dimethylcyclohexanediamine (25 mg, 0.17 mmol) was added and the reaction mixture was purged with argon for 15 min. The high pressure reaction vessel was sealed with a teflon screw cap and the reaction mixture was heated to 130 °C for 75 hours. The reaction mixture was then cooled to room temperature, diluted with DCM, and washed with aqueous NH_4OH (10 ml, 2 N). The organic phase was dried over $MgSO_4$, filtered, and

concentrated under reduced pressure. The resulting oil was purified with flash column chromatography on silica gel (10% DCM in hexanes) to afford **23** in 45% yield as a light yellow oil: ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.75 (dd, *J* = 7.2, 1.3 Hz, 2H), 6.67 (s, broad, 2H), 6.54 (dd, *J* = 7.2, 0.5 Hz, 2H), 5.37 – 5.25 (m, 2H), 3.48 (d, *J* = 7.7 Hz, 4H), 1.73 (s, 7H), 1.63 (s, 6H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 150.91, 148.68, 136.71, 130.01, 129.78, 119.83, 119.29, 117.67, 32.73, 25.83, 17.91; LRMS (FAB+): 352 $[M]^+$; HRMS calcd for $C_{22}H_{24}S_2$: 352.1319; found 352.1330.

Synthesis of the locked cyclic biphenylene 9a. Polyphosphoric acid (PPA, 1.0 ml) was added to an oven-dried Schenk tube with stirring bar and the Schlenk tube was backfilled with argon three times. Then, a solution of **23** (30 mg, 0.085 mmol) in toluene (3 ml, purged with argon for 15 minutes) was added to the Schlenk tube, the reaction mixture was heated to 80 °C and stirred vigorously for 16 hours. The reaction mixture was then cooled to room temperature and the toluene phase was decanted. The leftover PPA was hydrolyzed with H_2O (15 ml) and the resulting phosphoric acid solution was extracted with DCM. The combined organic phases were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The resulting oily residue was purified with flash column chromatography over silica gel (10% DCM in hexanes) to afford cyclic biphenylene **9a** in 76% yield as a single isomer and a yellow solid: Mp. 182–185 °C; ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.69 (s, 2H), 6.35 (s, 2H), 2.96 (s, broad, 4H), 1.97–1.86 (m, 4H), 1.25 (s, 12H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 147.65, 146.95, 140.52, 130.02, 116.49, 115.81, 38.33, 33.46, 30.22, 23.40; LRMS (FAB+): 352 $[M]^+$; HRMS calcd for $C_{22}H_{24}S_2$: 352.1319; found 352.1306.

About 2 mg of biphenylene **9a** were repurified by preparative HPLC (liquid phase: 20% H_2O in acetonitrile; flow rate: 10 ml/min; compound injected as a solution in 50 µL THF, retention time: 22 minutes). The repurified compound was used for the single molecule conductance measurements.

Synthesis of the locked cyclic fluorene 10a by Dr. Rachid Skouta (HPLC purification and characterization by me)**.** To a stirred solution of **24** (140 mg, 0.382 mmol) in toluene (10 mL), PPA (1 ml) was added at room temperature. The mixture was heated at 120 °C until the starting material was consumed (followed by ^{1}H NMR, after 24 hours). The reaction mixture was allowed to cool to 20 °C and poured into aqueous solution of sodium hydroxide. The organic layer was separated and the aqueous layer was extracted three times with DCM. The organic layers were dried over $MgSO_4$ and the solvent was evaporated under reduced pressure. The crude reaction mixture was purified by column chromatography (hexanes:EtOAc = 50:1) to provide an inseparable mixture of cyclic isomers **(10a-c)** in 75% yield.

About 2 mg of the mixture of isomers was separated via preparative HPLC (liquid phase: 20% H_2O in acetonitrile; flow rate: 10 ml/min; Compound injected as a solution in 50 µL THF, retention time: 24 minutes) to afford **10a** as a single isomer and a colorless solid: Mp. 173−175 C; ^{1}H NMR (300 MHz, CD_3OD) δ 7.84 (s, 2H), 7.19 (s, 2H), 3.72 (s, 2H), 3.16–2.91 (m, 4H), 2.10–1.86 (m, 4H), 1.47 (s, 12H); ^{13}C NMR (75 MHz, CD_3OD) δ 142.44, 142.04, 139.74, 131.07, 123.67, 118.25, 39.37, 36.29, 34.34, 30.80, 23.89; MS (FAB+): 366 $[M]^+$; HRMS calcd for $C_{23}H_{26}S_2$: 366.1476; found 366.1469.

Cyclic Voltametry (CV) Measurements. Electrochemistry was performed on a BAS CV-50W voltammetric analyzer with a three electrode cell. Each analyte was dissolved in acetonitrile containing 0.100 M tetrabutylammonium perchlorate (TBAP) as the supporting electrolyte to give 1.0 mM final concentration in analyte. Sample solutions were sparged with argon for 5 minutes prior to use. A platinum disc working electrode (BAS, d = 1.6 mm) and a platinum wire counter electrode were used. The reference electrode contained a silver wire with 10 mM silver nitrate in TBAP electrolyte solution. Cyclic voltammetry experiments were carried out at ambient temperature and were recorded at 100 mV/s. Potentials were swept from -500 mV to 1200 mV or from -2000 mV to 2000 mV. Ferrocene was used as an external calibrator and gave $E_{1/2}$ = 90 mV (vs $Ag/AgNO_3$) for each experiment.

3 Single molecule Conductance through pi-pi stacked benzene rings with direct carbon-Au links

3.1 An ideal electrical connection with direct Au-C contacts

In the past, gold break-junction based single molecule conductance measurements were always carried out with Lewis basic heteroatom gold binding groups such as amines, sulfides or phosphines.[22] While such binding groups provide stable binding geometries of the molecular wires with gold, they also introduce a very significant additional resistance to the molecular wire, which is also referred to as the Shottky barrier. For the first time, the Breslow and Venkataraman groups have now recently developed two alternative ways to link molecules to gold electrodes via direct Au-C contacts, thereby reducing the resistances of the Shottky barriers. While one approach is based on in situ chemical replacement of trimethyltin endgroups by gold[66] and will be described later in this dissertation, the other more direct approach employs the idea of using carbon-based π-systems as ligands to gold.[42]

The Nuckolls and Venkataraman groups found that unstrained pi-systems, such as polyenes or arenes do not bind to gold electrodes directly, without at least one Lewis basic heteroatom binding group.[67] However, it is well know that *strained* π-systems, such as e.g. present in [2.2]-paracyclophanes form stable complexes with many transition metals, silver being one of them.[68] Therefore, we decided to investigate whether simple [2.2]-paracyclophane would bind directly to gold electrodes in STM gold break junction single molecule conductance experiments. We purchased a bottle of [2.2]-paracyclophane (**1**) from Sigma Aldrich and Maria Kamenetska from the Venkataraman group tried out single molecule conductance experiments with it. We were pleased to see that [2.2]-

paracyclophane (**1**) did indeed bind to gold electrodes, showing a well defined conductance peak in the conductance histogram at ca. $1.0 \bullet 10^{-2}$ G_0 ().[42]

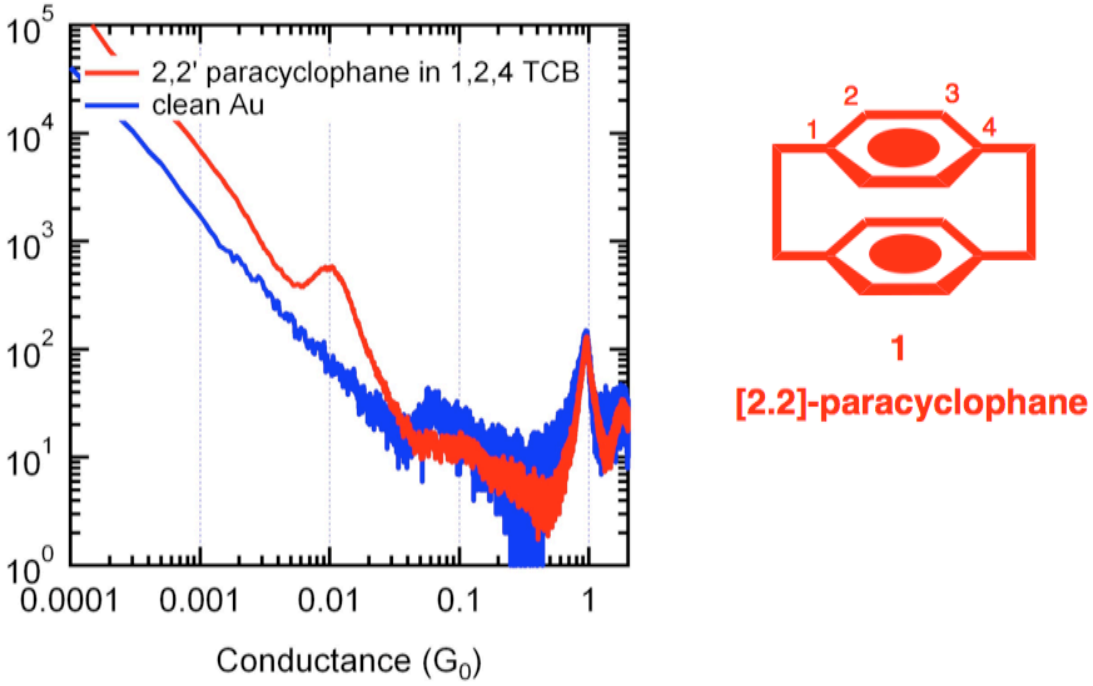


Figure 3.1. Single molecule conductance histogram of [2.2]-paracyclophane (**1**).[42]

A computational investigation of the binding of [2.2]-paracyclophanes to gold electrodes, which will be described in detail later, showed that the gold electrodes bind directly to the π system of the benzenoid rings in the cyclophanes, thereby contacting at most two carbon atoms at a time. The most preferred binding mode for a paracyclophane with four stacked benzenoid layers is illustrated schematically in Figure 3.2.

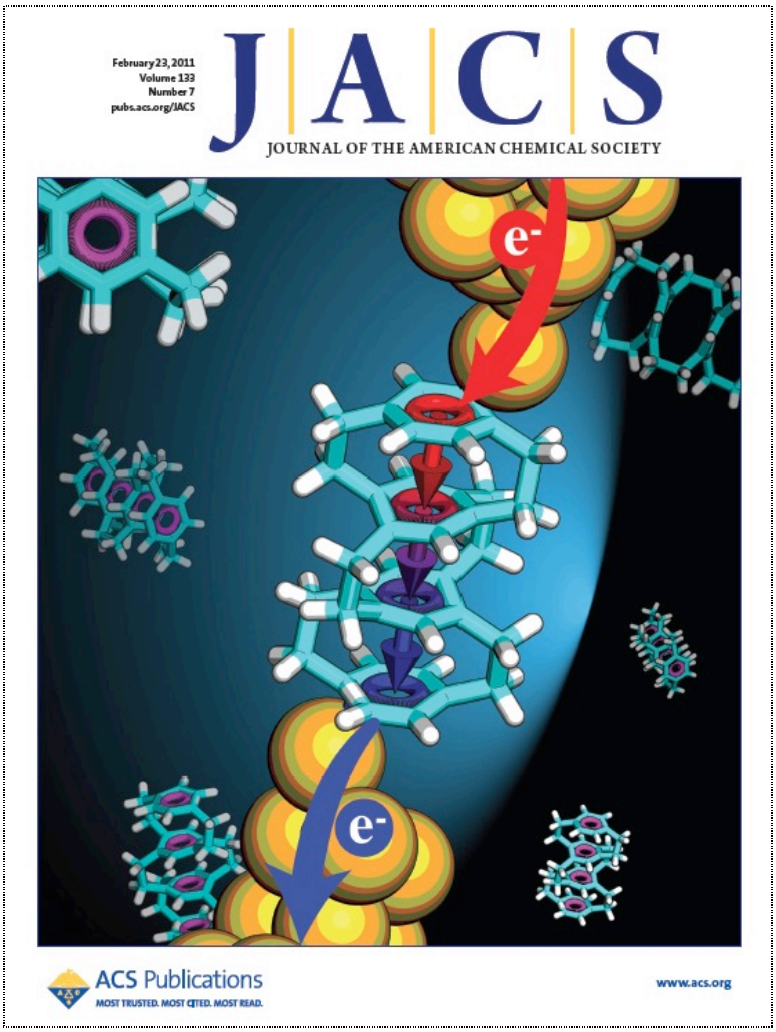


Figure 3.2. Cover Art[42] showing the direct Au-C electrical contacts achieved with [2.2]-paracyclophanes.

While our conductance measurements through a two-layer paracyclophane proved the concept that gold-carbon bonds can be employed to directly couple gold electrodes into π-systems, most applications of electron transport through stacked π-systems in nanotechnology involve multiple stacked arene rings. One such example where the details of electron transport through multiple stacked π-systems is critical are organic photovoltaic cells.[69] We soon realized that our newly discovered approach to directly couple strained arene rings to gold electrodes represents a unique handle to study the

details of electron transport through multiple stacked π-systems at the single molecule level, which will be illustrated in the following section.

3.2 Keep stacking in the third dimension

Besides greatly reducing the extra resistance of the Shottky barriers, another great advantage of our new approach to form direct Au-C bonds with the electrodes is that no gold binding groups need to be installed. This can simplify the syntheses of molecular wires very significantly. Therefore, since unsubstituted multiple stacked [2.2]-paracyclophes had been known for already several decades,[70-72] we decided to study the single molecule conductance of such paracyclophanes with multiple stacked benzene rings. Figure 3.3 summarizes our plans in this regard while the details of the single molecule conductance results we obtained with such systems will be discussed in the following sections.

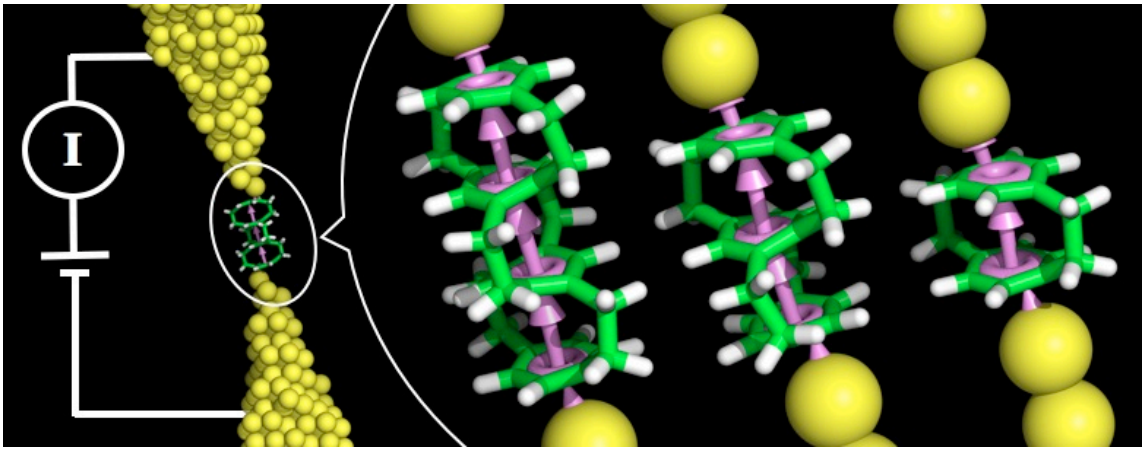


Figure 3.3. The single molecule electron transport through [2.2]-paracyclophanes with up to four stacked benzenoid rings can be studied with direct Au-C links.

3.3 Synthesis of paracyclophanes with multiple stacked benzene rings

The [2.2]-paracyclophanes **2** and **3** with up to four stacked benzenoid layers (Figure 3.4) were already known compounds[70-72] and I synthesized the triple stacked paracyclophane **2** following a modified version of the originally published procedure. My modified synthesis is shown in Scheme 3.1. A mixture of the quadruple stacked paracyclophanes **3a** and **3b** was synthesized by Dr. Zhanling Cheng, a former postdoc in the Breslow lab, following a similar procedure.[42]

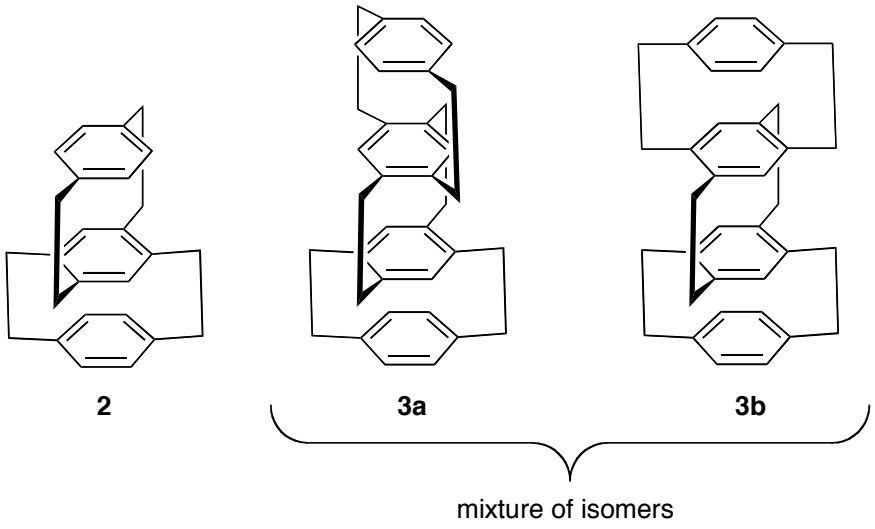
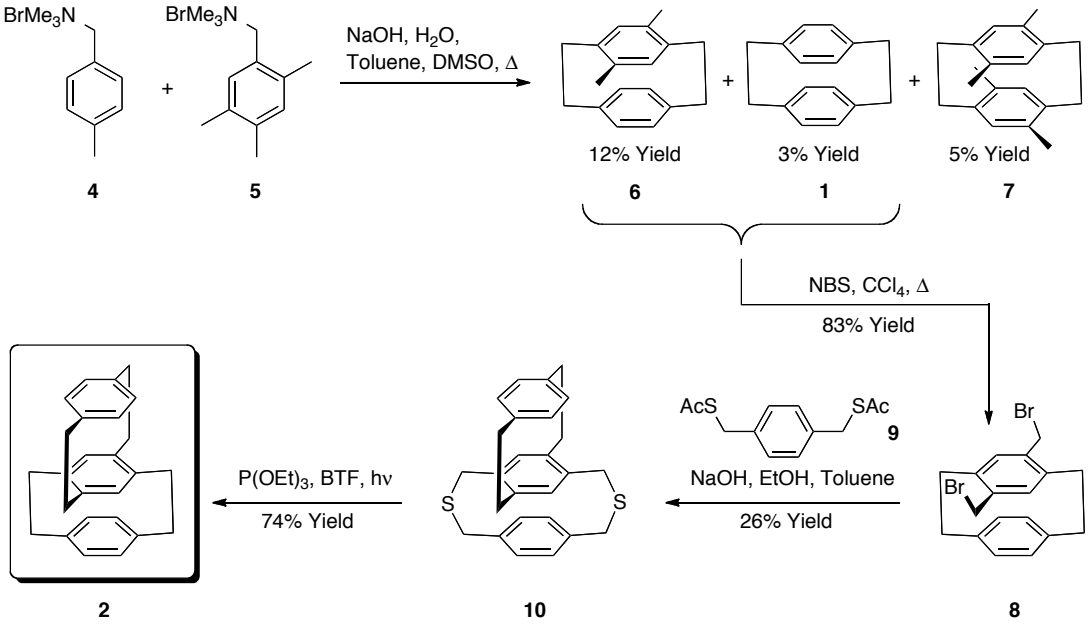


Figure 3.4. Paracyclophanes with multiple stacked benzenoid rings linked with saturated (ethylene) chains. The quadruple stacked paracyclophane **3** was isolated as a mixture (1.35:1.00) of the isomers **3a** and **3b**.



Scheme 3.1. The synthesis of the triple-stacked paracyclophane represents a modified version of the synthesis developed originally by Misumi *et. al*.[70-72] BTF is trifluoromethylbenzene.

In order to gain a better understanding of the binding of the paracyclophane compounds to gold electrodes, Dr. Rachid Skouta from the Breslow group also synthesized the methoxy substituted [2.2]-paracyclophane **12** and the tetrafluoro substituted [2.2]-paracyclophane **13** following known procedures.[42]

3.4 Single molecule conductance results

The single molecule conductance of the [2.2]-paracyclophanes **1** – **3** with multiple stacked benzene rings was determined with the STM gold break junction technique. All condutance measurements were carried out by Maria Kamenetska in the Venkataraman lab. Single-molecule junctions were created by repeatedly forming and breaking gold point contacts in a solution of the molecules in 1,2,4-trichlorobenzene, in a home-built

setup.[56] While compounds **2** and **3** were measured under Argon atmosphere, all other compounds were measured in air. For each molecule studied, the measured conductance traces reveal steps at molecule-dependent conductance values less than the quantum of conductance, $G_0 = 2e^2/h$; these are due to conduction through a molecule bonded in the gap between the two Au point contacts. Figure 3.5 shows conductance histograms generated (without any data selection) from over 5000 measured traces for compounds **1 – 3**.

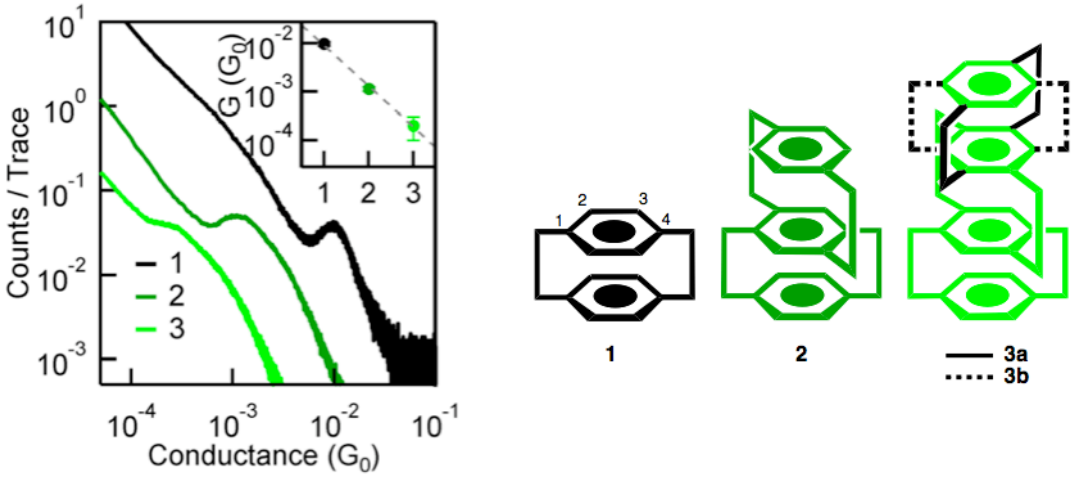


Figure 3.5. Conductance Histograms of compounds **1 – 3** generated by Maria Kamenetska in the Venkataraman lab using a linear bin size of 10^{-5} G_0 for compounds **1** and **2** and 10^{-6} G_0 for compound **3**. Inset: Conductance histogram peak versus number of paracyclophane units shown on a semi-log scale. The dotted line represents an exponential fit to the data with a decay constant β of (1.94 ± 0.25)/(stacked benzene unit). Error bars capture the variability in peak position. Compound **3** was measured as a mixture (1.35:1.00) of the isomers **3a** (with solid black bonds) and **3b** (with dashed black bonds), which might explain the slightly broader peak observed for compound **3**.

The inset in Figure 3.5a shows the position of the histogram peaks determined from Lorentzian fits to the data. We see that the molecular conductance decreases

exponentially with increasing number of stacked aromatic rings. This finding is consistent with a non-resonant tunneling mechanism of the electron transport through these π-stacked molecules, and strongly suggests that each gold electrode contacts the outermost benzene rings in compounds **1 – 3**, rather than the bridging ethylene groups. Further evidence for this conclusion is obtained by analyzing two-dimensional histograms[73] showing molecular conductance as a function of STM tip-sample displacement (Figure 3.6). The inset of Figure 3.6 shows that the most probable step length increases linearly with the number of stacked benzene rings, which can be explained only by concluding that the metal electrodes bind to the outermost benzene rings of the paracyclophanes.

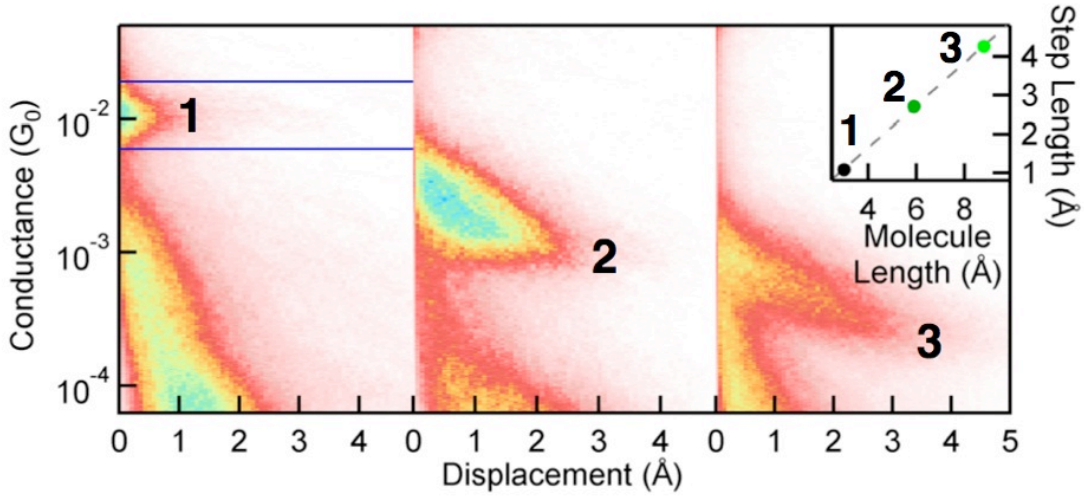


Figure 3.6. 2D-histograms showing molecular conductance as a function of STM tip-sample displacement for compounds **1 – 3** generated using a logarithmic binning with 10 bins/decade by Maria Kamenetska in the Venkataraman lab. The displacement dimension was binned linearly with a bin size of 0.072 Å. The tip-sample displacement is proportional to the length of the molecule in the break-junction.[74] The horizontal blue lines in the 2D histogram for compound **1** mark the section of the plot used to determine the step length. Inset: Measured step length plotted against the distance between the outermost benzene rings. The dashed line represents a linear least squares fit with slope ~0.5.

From an exponential fit to the conductance data shown in Figure 3.5a we were able to determine the decay constant β with $G \sim e^{-\beta N}$, as (1.94 ± 0.25)/(stacked benzene unit). Using a ring spacing of (3.07 ± 0.02) Å (Average ± standard deviation of all possible benzene centroid distances in compounds **1 – 3**, obtained from B3LYP/LACVP** structures of compounds **1 – 3** bound to two gold electrodes modeled as Au_2-clusters) between layers of stacked benzene rings this converts to a β of (0.63 ± 0.09)/Å. This is smaller than the value observed for alkanes, which show a beta of about 0.8/Å.[20] The significant difference between the decay constant of the π-π stacked paracyclophanes **1 – 3** and the one of alkyl chains indicates that the conductance channel in which electrons tunnel directly between the π-π stacked layers has indeed a major contribution to the overall conductance in [2.2]-paracyclophane type structures.

3.5 Strain is important for the binding to gold

It is well known that the strained benzene rings in [2.2]-paracyclophanes can form complexes with transition metals that are in general significantly more stable than the corresponding complexes involving normal, unstrained benzene rings.[75] Therefore, to better understand the role of strain in facilitating the binding of hydrocarbons to gold electrodes, Maria Kamenetska measured the unstrained [4.4]-paracyclophane (**11**, obtained from Sigma-Aldrich's rare chemical library). She found that **11** did not show any molecular conductance plateaus in traces, or conductance peak in the histogram to within the measurement limit of her experiment (Figure 3.7), indicating that **11** probably did not bind to gold electrodes. If compound **11** had bound to the gold electrodes, one would expect to see molecular conductance traces with low, but still detectable,

conductance as commonly observed for alkanes of similar length.[20] This result confirms that strain is indeed important to promote the direct gold-carbon contacts observed with [2.2]-paracyclophanes.

3.6 The electrical properties of substituted [2.2]-paracyclophanes

Further insight into the nature of the direct gold-electrode carbon contacts observed with paracyclophanes was obtained by studying substituted versions of [2.2]-paracyclophane (**1**). In general, electron-withdrawing substituents such as fluorines lower the Lewis-basicity of the paracyclophane benzene rings and therefore reduce the binding affinity for gold electrodes. On the other hand, electron-donating substituents such as methoxy should lead to stronger binding to the electrodes. In order to test this hypothesis Dr. Rachid Skouta synthesized fluorine and methoxy substituted versions of [2.2]-paracyclophane (compounds **12** and **13**) following known procedures.[42] While we found that compound **13** with four fluorine substituents did not bind to gold electrodes, results with the methoxy substituted compound **12** were very similar to those of the unsubstituted [2.2]-paracyclophane (**1**, Figure 3.7).

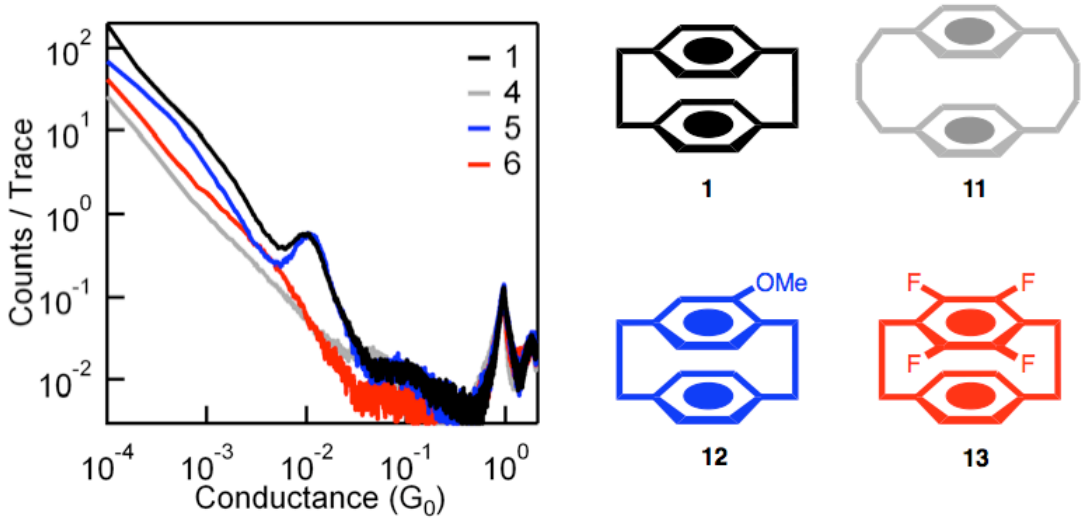


Figure 3.7. The effects of strain and substitution on the single molecule conductance of [2.2]-paracyclophanes. While the commercial [4.4]-paracyclophane (**11**) does not show any binding to the gold electrodes due to lack of strain in the system, the tetrafluoro substituted paracyclophane **13** (synthesized by Dr. Rachid Skouta) also shows no binding to the gold. However, the methoxy substituted compound **12** (synthesized by Dr. Skouta as well) binds to gold and shows a conductance indistinguishable from the unsubstituted cyclophane **1**. All conductance histograms were generated using a linear bin size of 10^{-4} G_0 by Maria Kamenetska in the Venkataraman lab.

Further insight into how our paracyclophane molecular wires directly contact gold electrodes and conduct electricity was finally obtained with DFT calculations, which are described in detail in the following section.

3.7 Computational investigation the paracyclophanes-Au wires

To understand the relative gold binding affinities of compounds **1** – **3** and **11** – **13** in a more quantitative fashion, I carried out DFT based calculations of the molecular binding energy to gold dimers, making the assumption that a linear chain of gold atoms links the molecule to the electrode. All binding energy calculations were performed at the spin unrestricted B3LYP/LACV3P**++//B3LYP/LACVP** level with the Jaguar software package. The pseudospectral method was employed to speed up the SCF cycles.

Default grids and convergence criteria were employed. It is important to note here that binding energies calculated with gold dimers cannot be compared directly with those of single gold atoms due to an odd/even effect on the binding energy.[76] However, our initial focus was on the trends in calculated binding energies, which are robust. Calculations with larger gold clusters, which will be necessary to get more accurate gold binding energies of the compounds studied, are currently in progress and will be reported elsewhere.

I found two stable binding geometries for the gold dimers bound to compound **1**. The first has the gold coordinating the C_1-C_2 "double" bond in an η^2-fashion, and the second with the gold attached in η^1-fashion to C_1 of [2.2]-paracyclophane (with C_1 and C_2 defined in Figure 3.5). The calculated gold binding affinities of 0.74 eV (17 kcal/mole) are very similar to the one reported for C-60 (0.67 eV) with a gold dimer electrode model.[77]

To compare the gold dimer binding energies of the cyclophane compounds **1 – 3** and **11 – 13** to the ones of traditional Lewis-basic heteroatom link groups, we also calculated the gold dimer binding affinity of methylamine (**14**) and dimethylsulfide (**15**), which are known to bind to undercoordinated Au in the experiments. and dimethylether (**16**) that does not bind to Au in the experiments. For the amine-terminated molecule (**14**), the calculated gold dimer binding energy is significantly larger than the ones reported with either a single gold atom or larger gold clusters.[20] Nonetheless, these gold dimer binding affinities serve as a guide for the relative gold binding strengths of the various compounds studied here.

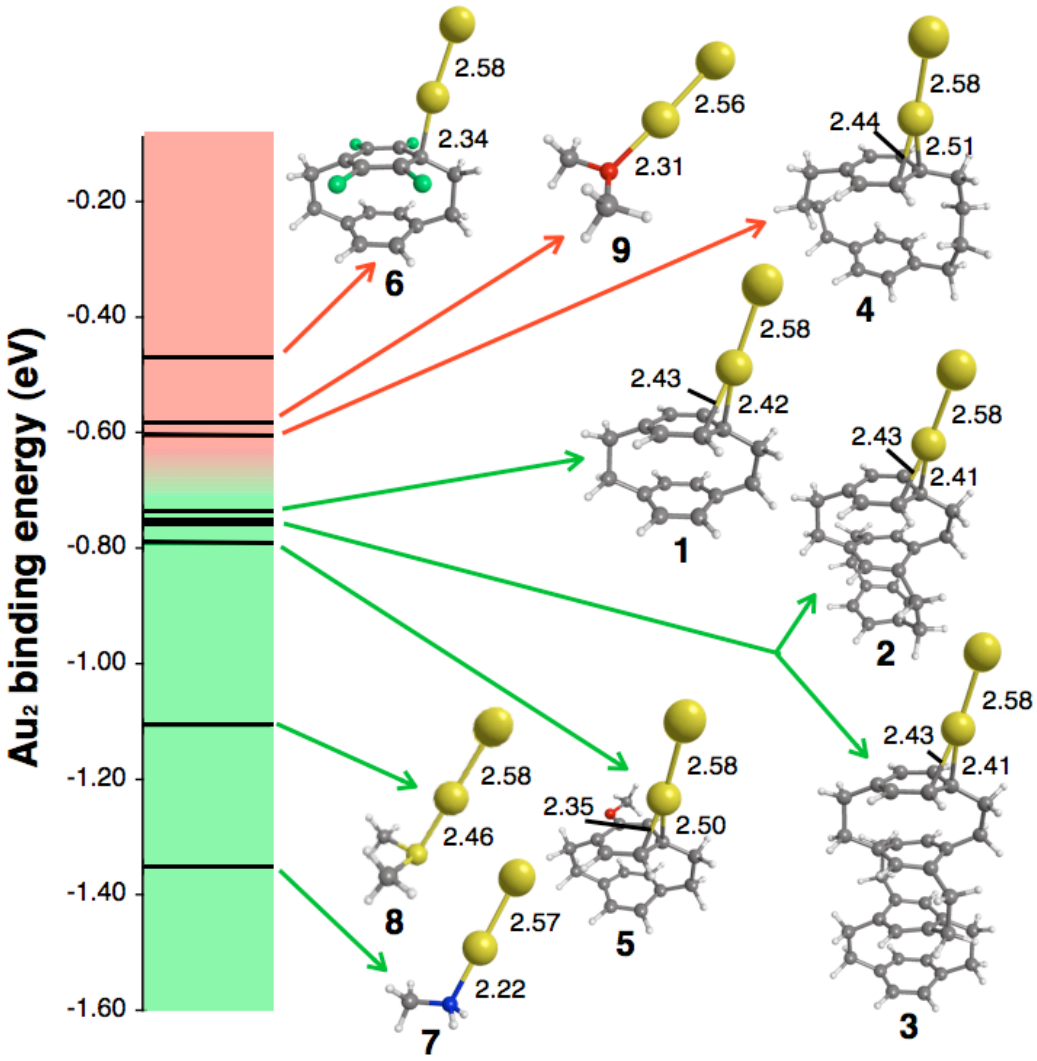


Figure 3.8. Optimized structures (B3LYP/LACVP** level) of compounds **1 – 3** and **11 – 16** bound to gold dimers (bond distances in Å). The most stable binding configurations are shown for all compounds. The green gold dimer binding energy range represents compounds that successfully bind to gold electrodes in STM break-junction conductance measurements.

Figure 3.8 shows the calculated gold dimer binding energies for compounds **1 – 3** and **11 – 16**. We see that all molecules with gold dimer binding energies above 0.7 eV actually bind to gold electrodes in STM break-junction measurements, allowing conductance measurements. Compounds with lower gold dimer binding energies do not show any peaks in conductance histograms, consistent with the understanding that these do not bind to undercoordinated Au on the electrodes. Finally, these calculations show

that paracyclophanes bind through their outer π system to the C_1 carbon or to the C_1-C_2 bond.

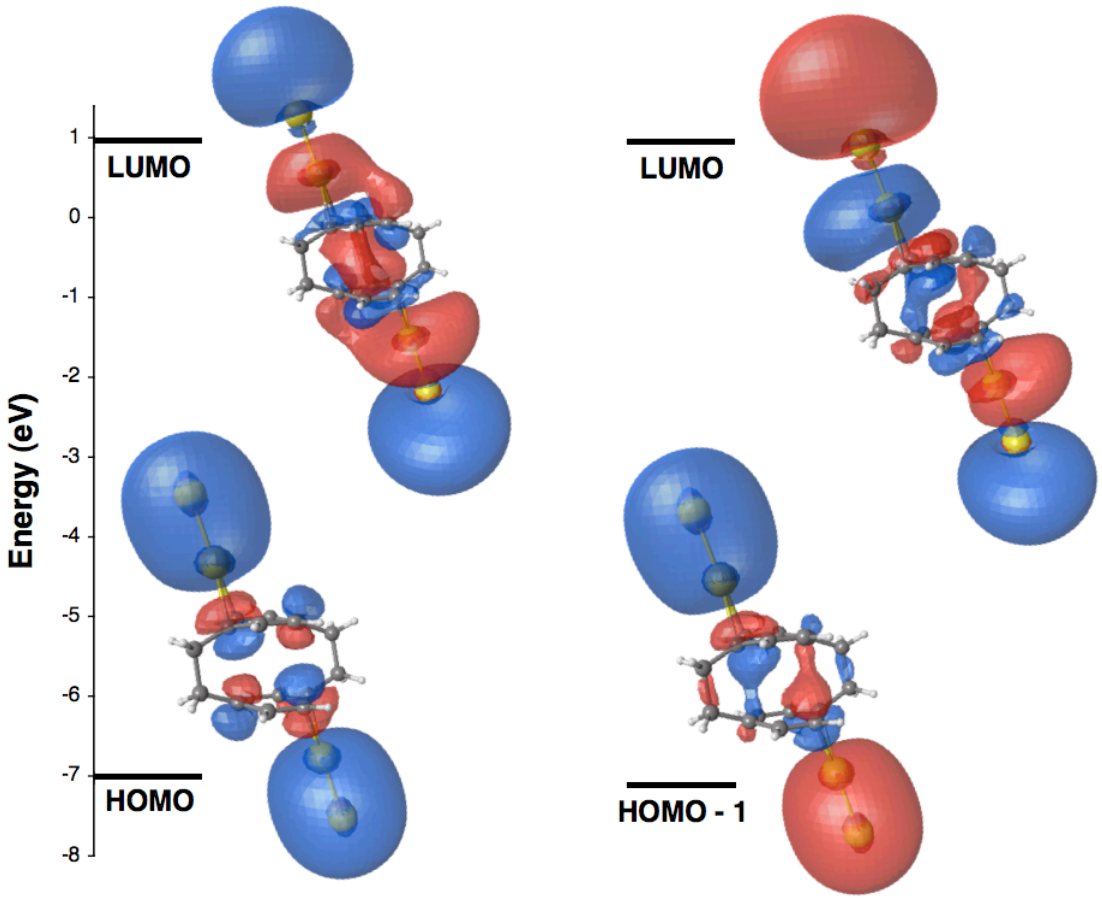


Figure 3.9. The HF frontier molecular orbitals of [2.2]-paracyclophane (1) bound to two gold dimers indicate that the major conductance channels indeed involve tunneling of the electrons directly from one-pi system to the other. Isosurface plots (contour value = 0.01) are shown for all frontier molecular orbitals. The two LUMOs shown are degenerate. Orbitals with bonding character between the gold dimers and the benzene rings are lower in energy and are not shown in this plot.

Frontier molecular orbitals determined by analyzing the molecular Hartree-Fock orbitals (at the RHF/LACVP** //UB3LYP/LACVP** level, computed with Gamess[78]) are shown in Figure 3.9 for molecule **1** bound to two gold dimers. We see that the frontier molecular

orbitals are mainly localized on the benzene rings and on the gold atoms, indicating that the Au is indeed coupled to the molecular π system. This provides further evidence that during electron transport through the π-π stacked systems studied herein transport is through the π system, rather than the σ-type molecular orbitals of the ethylene link groups.

3.8 Single molecule conductance through unstrained stacked aromatic rings

While our single molecule conductance results obtained with the ethylene bridged multilayered paracyclophanes demonstrated that strained arene rings can indeed form direct Au-C links with electrodes, these results also provided an initial idea about the conductance behavior of stacked aromatic molecules. However, since in many real applications, conductance through *unstrained* stacked pi-systems is important, the question arises of how well the *strained* multi layered cyclophanes we studied initially actually represent the unstrained stacked systems of importance to many real applications. To address this question, we are currently investigating the single molecule conductance through [3.3]-paracyclophanes, which are much less strained and therefore more closely resemble practically useful pi-stacked systems. My DFT calculations predict a gold dimer binding energy for simple [3.3]-paracyclophane, which is very close to the value predicted for [4.4]-paracyclophane (**11**). Therefore, it is very likely that, as observed for [4.4]-paracyclophane, [3.3]-paracyclophane will not bind directly to gold electrodes during the conductance measurements neither, which makes it necessary to attach gold binding groups to the system. While attaching gold binding groups in specific positions increases the synthetic challenge of producing the compounds, it will also allow

to study symmetry effects on molecular conductance in pi-stacked systems, which have been predicted by Professor Ratner and coworkers to be opposite as in the case of normal arene rings.[36,37] Therefore, studies of unstrained [3.3]-cyclophanes with gold binding groups in specific positions are currently under way in the Breslow and Venkataraman groups and will be reported in due time.

Furthermore, Alice Chang and Christine Yeh, two undergraduate students in the Breslow lab, started investigating the possibility of using strained alkenes as gold binding groups, which will help to generalize the concept of forming direct Au-C bonds with pi-systems. Results from this project will also be reported in greater detail in the future.

3.9 Experimental details

General procedures. *Chemicals:* All solvents, inorganic salts, and organic reagents were purchased from commercial sources and used without further purification unless otherwise mentioned. A mixture of compounds **3a** and **3b** was prepared by Dr. Zhanling Cheng, while Dr. Rachid Skouta synthesized compounds **12** and **13** according to known procedures.[42] Compounds **4**[79] and **5**[80] were prepared following known procedures by me.

Chromatography: Merck pre-coated 0.25 mm silica plates containing a 254 nm fluorescence indicator were used for analytical thin-layer chromatography. Flash chromatography was performed on 230-400 mesh silica (SiliaFlash® P60) from Silicycle.

Spectroscopy: NMR spectra were obtained on a Bruker DPX 300 or 400 MHz spectrometer. Spectra were analyzed with the MestreNova Software (Version 6.1). Chemical shifts for the proton spectra are reported in ppm (δ) relative to chloroform (δ

7.26). Chemical shifts for the carbon spectra are reported in ppm (δ) relative to chloroform (δ 77.16) as well. CI-MS spectra were taken on a Nermag R-10-10 instrument. FAB MS spectra were taken on a JEOL JMS-DX-303 HF instrument using either glycerol or p-nitrobenzyl alcohol as matrices. Matrix assisted laser desorption ionization (MALDI) mass spectra were acquired using an Applied Biosystems Voyager DE Pro time-of-flight mass spectrometer. Positive ion mass spectra were acquired in the linear mode using a nitrogen laser (337nm). Instrument settings were as follows: accelerating voltage, 21,000 volts; grid voltage, 95%; guide wire, 0.05%; extraction delay time, 200 nsec. All data processing was performed using Applied Biosystems Data Explorer v 4.0.0.0.

Synthesis of paracyclophane 6.[81] A modified literature procedure[82] for the synthesis of unsubstituted [2.2]-paracyclophane was followed: To a solution of NaOH (6.0 g, 0.15 mol) in 9 ml H_2O was added a mixture of the ammonium salts **5** (2.79 g, 10.25 mmol) and **4** (2.50 g, 10.25 mmol), which had been mixed and pulverized in a mortar before the addition. Then, DMSO (16 ml) and toluene (80 ml) were added as well and the reaction mixture was heated at 100º C for 55 hours. The reaction mixture was then cooled to room temperature, the organic phase was separated and the aqueous phase extracted three times with 20 ml toluene. The combined organic phases were then dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was then purified with flash column chromatography (gradient elution from pure hexanes to 50% DCM in hexanes) over silica gel to give a 61:14:25 mol% mixture of compounds **6**, **7** and **1** in 20% yield. The mixture was then recrystallized from MeOH (heated above the boiling point to about 90º C in a closed pressure vessel) to give colorless crystals, which

were a 84:14 mol% mixture of compounds **6** and **1**, while compound **7** stayed in the MeOH solution. Since compound **1** does not have a free methyl group, it can not interfere with the following bromination step. Therefore, we used the mixture of compounds **6** and **1** without further purification for the bromination step and were able to separate the compounds after bromination as described below. Our spectroscopy data for compound **6** matched the values reported in the literature.[81] ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.82 (dd, *J* = 7.9, 1.9 Hz, 2H), 6.39 (dd, *J* = 7.9, 1.8 Hz, 2H), 6.03 (s, 2H), 3.27 (ddd, *J* = 13.1, 9.9, 1.8 Hz, 2H), 3.17 – 2.92 (m, 4H), 2.66 (ddd, *J* = 13.3, 10.5, 6.6 Hz, 2H), 2.11 (s, 6H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.45, 137.99, 136.43, 134.43, 132.50, 128.60, 33.65, 33.18, 19.76; LRMS (EI+): 236 $[M]^+$; HRMS calcd for $C_{18}H_{20}$: 236.1565; found 236.1567.

Synthesis of the paracyclophane dibromide 8.[81] A modified literature procedure was followed for the synthesis of **8**.[81] The inseparable mixture of 86 mol% dimethylparacyclophane **6** and 14 mol% [2.2]-paracyclophane (**1**) obtained in the previously described step of our synthesis (0.250 g, 0.920 mmol of **6**) was dissolved in 10 ml CCl_4. Then, NBS (0.328 g, 1.84 mmol) was added and the reaction mixture was refluxed at 80º C for 3 hours. The reaction mixture was then cooled to room temperature and the succinimide side product was filtered off. After washing the filter cake with CCl_4, the filtrate was concentrated in vacuo and the residue was purified with flash column chromatography (10% DCM in hexanes) over silica gel to give **8** as a colorless solid in 83% yield. Our spectroscopic characterization matches the one reported in the literature.[81] ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.63 (dd, *J* = 7.9, 1.3 Hz, 2H), 6.42 (dd, *J* = 7.9, 1.3 Hz, 2H), 6.36 (s, 2H), 4.49 (d, *J* = 10.4 Hz, 2H), 4.24 (d, *J* = 10.4 Hz, 2H), 3.42 (ddd, *J* = 10.3, 5.8, 0.8 Hz, 2H), 3.19 (ddd, *J* = 10.3, 5.8, 0.8 Hz, 2H), 3.04 (ddd, *J* = 13.3,

10.1, 6.1 Hz, 2H), 2.85 (ddd, *J* = 13.6, 10.8, 6.1 Hz, 2H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 139.33, 138.94, 137.29, 132.59, 130.26, 34.25, 32.56, 32.49.

Synthesis of the triple layered cyclophane 10. A modified literature procedure was followed for this reaction.[81] A solution of the paracyclophane dibromide **8** (17 mg, 0.043 mmol) and the *p*-xylene dithioester **9** (8 mg, 0.043 mmol) in 10 ml toluene was carefully degased by bubbling a stream of argon gas through the solution for ca. 15 minutes. Then, the solution containing **8** and **9** was added to a solution of KOH (100 mg, 1.78 mmol, also degased with argon for ca. 15 minutes) in ca. 90% aqueous EtOH at room temperature. The addition was carried out with a syringe pump at an addition rate of 0.5 ml/hour. After an additional 12 hours of stirring at room temperature under argon atmosphere, the reaction mixture was concentrated under reduced pressure, the residue was taken up in methylene chloride, washed with H_2O, dried over $MgSO_4$, filtered and evaporated to dryness *in vacuo*. The crude product was then purified with preparative TLC (25% DCM in hexanes) with a TLC plate containing a 1000 μm thick layer of silica gel to give the triple stacked cyclophane **10** as a colorless solid in 26% yield. Our spectroscopic characterization matched the one reported in the literature.[81] ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.70 (dd, *J* = 7.9, 1.6 Hz, 2H), 6.62 (dd, *J* = 7.9, 1.6 Hz, 2H), 6.40 (dd, *J* = 7.9, 1.8 Hz, 2H), 6.21 (dd, *J* = 7.9, 1.7 Hz, 2H), 5.79 (s, 2H), 3.80 (d, *J* = 14.7 Hz, 2H), 3.74 – 3.60 (m, 4H), 3.47 (d, *J* = 14.7 Hz, 2H), 3.30 (ddd, *J* = 13.3, 10.0, 1.3 Hz, 2H), 3.05 – 2.94 (m, 2H), 2.77 (ddd, *J* = 13.3, 9.9, 6.9 Hz, 2H), 2.55 (ddd, *J* = 13.6, 10.4, 6.9 Hz, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.37, 137.30, 136.54, 135.47, 134.77, 132.33,

129.79, 129.62, 129.33, 38.40, 36.41, 34.23, 32.78; LRMS (FAB+): 402 $[M]^+$; HRMS calcd for $C_{26}H_{26}S_2$: 402.1476; found 402.1486.

Synthesis of the triple layered cyclophane 2. A modified literature procedure was followed for this reaction.[81] To a solution of the cyclophane **10** (16 mg, 0.040 mmol) in 13 ml anhydrous benzotrifluoride in a quartz test tube closed with a septum was added $P(OEt)_3$ and the reaction mixture was purged with a stream of argon gas for ca. 15 minutes. Then, the reaction mixture was irradiated in a reactor with high pressure mercury lamps for one hour. After the irradiation, the reaction mixture was concentrated under reduced pressure and the product was purified with preparative TLC (hexanes) with a TLC plate containing a 1000 μm thick layer of silica gel to give the triple stacked cyclophane **2** as a colorless solid in 74% yield. Our spectroscopic characterization matched the one reported in the literature.[81] ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.21 (s, 8H), 5.42 (s, 2H), 3.11 (t, *J* = 12.4 Hz, 4H), 2.96 (t, *J* = 11.6 Hz, 4H), 2.72 (ddd, *J* = 13.0, 10.0, 6.6 Hz, 4H), 2.53 (ddd, *J* = 12.9, 10.2, 6.6 Hz, 4H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 140.22, 139.14, 135.64, 132.87, 131.95, 35.65, 32.50; LRMS (FAB+): 338 $[M]^+$; HRMS calcd for $C_{26}H_{26}$: 338.2035; found 338.2050.

4 Formation of in situ Au-C electrical contacts from tin based precursors

4.1 Highly conducting Au-C links from $SnMe_3$ precursors

Around the same time when the direct Au-C links with the strained paracyclophanes were discovered, my lab mate in the Breslow group at that time, Dr. Zhanling Cheng, was investigating other elements of the periodic table as potential gold binding groups. Dr. Cheng had already tried trimethylsilyl groups as potential contacts to gold electrodes, but trimethylsilyl groups did not bind to gold in the STM break junction conductance experiments. As going down along one column in the periodic table had proven a successful strategy in the past, he decided to try trimethyltin terminated molecules. He synthesized 1,4-bis(trimethylstannyl)benzene[66] (**1**) and to our surprise Jonathan Widawsky from the Venkataraman lab was actually able to obtain molecular conductance traces with that molecule. Molecule **1** conducted significantly higher than 1,4-benzenediamine, which represented a very exciting result.

In order to see whether the tin endgroups could be used in general to contact molecules to gold electrodes, Professor Venkataraman suggested to also try measurements with trimethyltin terminated alkanes. Therefore Dr. Rachid Skouta and I synthesized a series of such trimethyltin terminated alkanes, following known procedures.[81] To our positive surprise, all of those molecules bound to the gold electrodes and showed very high conductance, ca. 100 times larger than the conductance values observed for the corresponding diamine terminated alkanes. Furthermore, as shown in Figure 4.1, the conductance values decayed exponentially with increasing alkyl chain length, with a decay constant which perfectly matched the one observed for diamine

terminated alkanes. These results indicated that we had indeed found a general new endgroup based on tin, which was leading to very high conductance.

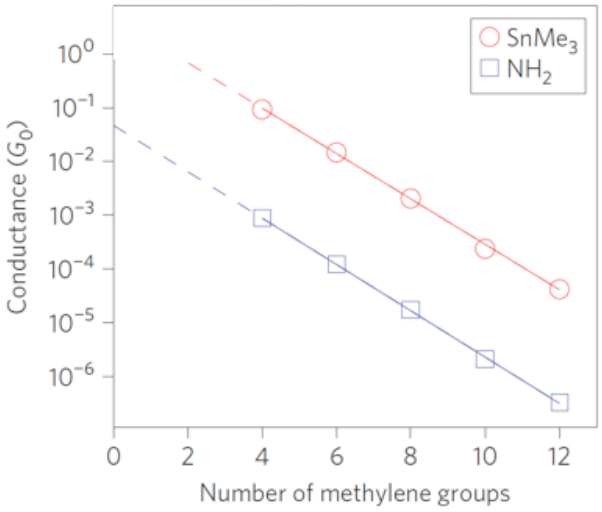


Figure 4.1. In the polymethylene series, the trimethyltin derivatives (red) are much better electrical conductors than are those with amino groups (blue).[81]

At that point, we had no idea how the trimethyltin terminated molecule contacted the gold electrodes and were wondering what could be the origin of the very high conductance values observed. I therefore carried out preliminary DFT (B3LYP/LACVP**) optimizations of the 1,4-bis(trimethyltin)benzene (**1**) with one Au atom in proximity of each Sn to see whether there was any chance that the Sn would directly contact the Au electrodes. I started the DFT optimizations from different starting geometries and while in most of the optimizations the Au simply moved far away from the Sn, indicating that there was absolutely no attracting interaction, in one optimization the Au atom inserted itself into one of the Sn-C bonds. This result suggested that

insertion of gold into a C-Sn bond should be a process with relatively low barrier and was one of our first indications that tin compounds could react with gold electrodes in situ to form direct Au-C bonds. During our next meeting, Jonathan Widawsky then told us that the 1,4-Bis(trimethylstannyl)benzene (**1**) actually did not show conductance peaks instantaneously, but that conductance peaks actually did not appear until ca. 1 hour after the start of the measurements. This was a further indication that a chemical reaction must have happened at the beginning of the measurement before the 1,4-bis(trimethylstannyl)benzene (**1**) finally started to conduct electricity.

In addition to the main conductance peaks of the trimethyltin terminated polymethylene chains, Jonathan Widawsky also observed conductance peaks corresponding to alkyl chains of double length. While we originally suspected that these additional peaks could stem from trace impurities potentially formed during the synthesis of the compounds, we soon realized that a gold insertion mechanism into the tin-carbon bonds could actually explain these additional peaks as well. We now believe that dimers and even trimers of the original trimethyltin terminated molecules are formed via in situ coupling reactions taking place on the gold surface, after the gold has inserted into the tin carbon bonds.

All of the above mentioned experimental findings are consistent with the postulated insertion mechanism of the gold electrodes into the Sn-C bonds, however, much more conclusive evidence for such a mechanism was obtained by comparing the conductance of the trimethyltin terminated molecules to molecular wires which contain a synthetic Au-C bond by design. These experiments, which were initiated by Dr. Zhanling Cheng

and finalized by Dr. Wenbo Chen, two postdocs in the Breslow lab, are summarized in the following section.

4.2 Molecular wires with synthetic Au-C bonds

Searching for ways to prove the postulated insertion mechanism of the Au electrodes into the trimethyltin terminated molecular wires, it occured to us that molecular wires with synthetic Au-C bonds might provide strong evidence for our postulated mechanism if the conductance of the molecular wires with synthetic Au-C bonds would be approximately the same as the conductance of the trimethyl tin terminated wires. This concept is illustrated in Figure 4.2. Dr. Zhanling Cheng therefore synthesized 1,4-bis(triphenylphosphinegold)benzene (**2**), a molecular wire with two gold atoms attached to the para positions of benzene by design. Compound **2** indeed showed a single molecule conductance, which was indistinguishable from the one observed for 1,4-bis(trimethylstannyl)benzene (**1**), which provided strong evidence for our postulated gold insertion mechanism. A little later, Dr. Wenbo Chen was able to successfully synthesize analogous triphenylphosphinegold terminated alkyl chains as well. In full agreement with our postulated mechanism, these gold terminated alkyl chains also showed the same conductance values as the corresponding tin compounds.[81]

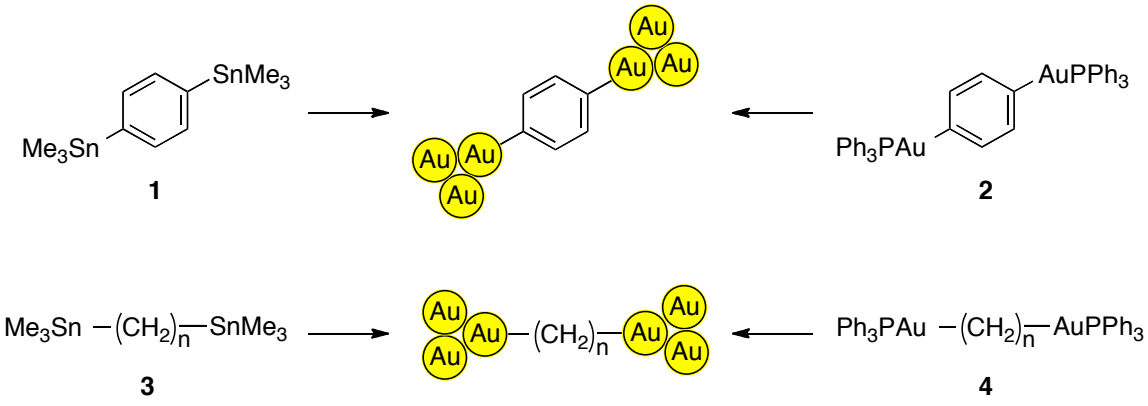


Figure 4.2. The trimethyltin derivatives and the gold/phosphine derivatives become directly attached to the gold electrodes, which leads to identical single molecule conductance traces for both systems.[60]

Apart from the highly convincing experimental evidence mentioned above, Dr. Hector Vazquez and Dr. Mark Hybertsen also performed DFT calculations of the electron transmission probabilities across molecular junctions with direct Au-C contacts, results of which are in good agreement with the experimental observations mentioned above as well.[81] Therefore, we are confident that the gold electrodes indeed contact the newly discovered trimethyl tin based molecular wires by inserting into the Sn-C bonds.

4.3 Au-C bonds that couple into conjugated pi-systems

As mentioned previously, our trimethyltin terminated alkane chains conduct about 100 times higher than the corresponding diamine terminated polymethylene chains. However, even though the tin terminated alkanes show the highest conductance values ever observed for such systems up to date, the tin endgroups do not solve the problem of the relatively fast exponential conductance decay with increasing chain length of saturated molecules. This very fast exponential decay of the conductance intrinsic to saturated chains limits the use of saturated systems for molecular electronics and it is

therefore highly desirable to attach our tin terminated endgroups to molecular wires with delocalized π-systems, which show a much slower decay of conductance with increasing length.

We have found that Au electrodes attached directly to aromatic rings (Figure 4.3a) couple mainly into the sigma system of the conjugated wires. Such an arrangement is for example formed with 1,4-bis(trimethylstannyl)benzene (**1**). Therefore in order to couple the Au-C bonds efficiently into a conjugated π-system, the gold will need to be attached one carbon further away from the conjugated system, in a benzylic or allylic type arrangement as shown in Figure 4.3b.

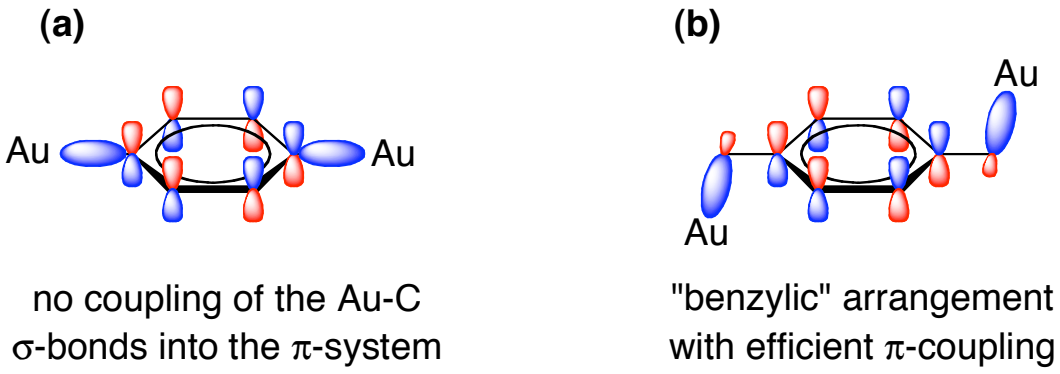


Figure 4.3. How to achieve efficient coupling of Au-C σ-bonds with the conjugated π-systems of molecular wires, schematically illustrated as the HOMO of benzene. **(a)** Direct attachement of the Au does not lead to efficient π-coupling as the Au-C σ-bond is orthogonal to the π-system. **(b)** Attachement of the Au in the benzylic position should lead to efficient π-coupling.

Initial results along these lines obtained by Dr. Wenbo Chen and Jonathan Widawski show very high conductance for p-xylene (**5**) and 4,4'-dimethylbiphenyl (**6**) molecular wires terminated with trimethyltin groups, again ca. 100 times higher than the corresponding wires with amine link groups. Detailed results about this very new and exciting study will be reported in the literature in due time.

5 Concluding remarks

While in the first part of this dissertation described some of my work towards new computational methodologies for predicting selectivities of chemical reactions, the second part focuses on the design, the synthesis and the electrical properties of novel molecular wires that might advance the development of future functional single molecular devices. We were able to synthesize the first air-stable partially antiaromatic single molecular wires and study their bulk and single molecule electrical properties with Cyclic Voltammetry and STM break junction conductance measurements. Our conductance measurements showed that the antiaromaticity of the biphenylene molecular wires does not significantly increase their conductance relative to fluorenyl molecular wires with "non-aromatic" cyclopentadiene cores. However, we also find that a significant degree of antiaromaticity is indeed present in the central ring of biphenylene, as demonstrated by the unusual aromatic substitution pattern, the distorted structure, and the easy oxidation of the system.

Apart from the first successful studies of partially antiaromatic molecular wires, my Ph. D. work has also contributed to the discovery of two novel ways to connect single molecules to gold electrodes with direct Au-C contacts. In the first such approach, strained arene rings in [2.2]-paracyclophanes were found to directly contact gold electrodes, thereby making direct contacts between the π-electrons of the aromatic systems and the gold electrodes. This new Au-C contact approach made it possible for us to study, for the first time, conductance through multiple stacked arene rings at the single molecule level with well defined gold break junction measurements. We are currently continuing to explore single molecule conductance of unstrained cyclophane systems, as

such systems are of even greater practical interest than the strained [2.2]-paracyclophanes we already studied. In addition, the Ratner and Wasiliewski Research groups have recently predicted interesting symmetry effects, which should be present for conductance through unstrained stacked arene rings,[36,37] which we are currently also trying to explore experimentally with our unstrained paracyclophane systems. Results of these studies will be reported in due time.

The second approach for making direct Au-C contacts recently found by the Breslow and Venkataraman labs involves the use of trimethyltin terminated "linker" groups, which get replaced by the gold electrodes in situ during the conductance measurements and show very high conductance (up to 100 times higher than with conventional linker groups). Very recent results show that is also possible to use this newly developed linker methodology to get very high conductance (again up to 100 times higher than with conventional amine groups) through π-conjugated molecular wires. We are continuing to further investigate how this new, exciting linker technology affects the conductance of molecular wires and will hopefully be able to report even more exciting results soon.

6 References

(1) Moore, G. *Electronics* **1965**, *38*, 114.

(2) Carroll, R. L.; Gorman, C. B. *Angew. Chem., Int. Ed.* **2002**, *41*, 4379.

(3) Joachim, C.; Gimzewski, J. K.; Aviram, A. *Nature* **2000**, *408*, 541.

(4) Nitzan, A.; Ratner, M. A. *Science* **2003**, *300*, 1384.

(5) Aviram, A.; Ratner, M. A. *Chem. Phys. Lett.* **1974**, *29*, 277.

(6) Robertson, N.; McGowan, C. A. *Chem. Soc. Rev.* **2003**, *32*, 96.

(7) James, D. K.; Tour, J. M. *Molecular Wires: From Design to Properties*; DeCola, L., Ed. 2005; Vol. 257, p 33.

(8) Li, C.; Pobelov, I.; Wandlowski, T.; Bagrets, A.; Arnold, A.; Evers, F. *J. Am. Chem. Soc.* **2008**, *130*, 318.

(9) Venkataraman, L.; Klare, J. E.; Nuckolls, C.; Hybertsen, M. S.; Steigerwald, M. L. *Nature* **2006**, *442*, 904.

(10) Xu, B. Q.; Tao, N. J. J. *Science* **2003**, *301*, 1221.

(11) Whalley, A. C.; Steigerwald, M. L.; Guo, X.; Nuckolls, C. *J. Am. Chem. Soc.* **2007**, *129*, 12590.

(12) Guo, X. F.; Small, J. P.; Klare, J. E.; Wang, Y. L.; Purewal, M. S.; Tam, I. W.; Hong, B. H.; Caldwell, R.; Huang, L. M.; O'Brien, S.; Yan, J. M.; Breslow, R.; Wind, S. J.; Hone, J.; Kim, P.; Nuckolls, C. *Science* **2006**, *311*, 356.

(13) Tam, I. W.; Yan, J. M.; Breslow, R. *Org. Lett.* **2006**, *8*, 183.

(14) Kushmerick, J. G.; Naciri, J.; Yang, J. C.; Shashidhar, R. *Nano Lett.* **2003**, *3*, 897.

(15) Blum, A. S.; Yang, J. C.; Shashidhar, R.; Ratna, B. *Appl. Phys. Lett.* **2003**, *82*, 3322.

(16) Wu, S. M.; Gonzalez, M. T.; Huber, R.; Grunder, S.; Mayor, M.; Schonenberger, C.; Calame, M. *Nat. Nanotechnol.* **2008**, *3*, 569.

(17) Kim, B.; Beebe, J. M.; Jun, Y.; Zhu, X. Y.; Frisbie, C. D. *J. Am. Chem. Soc.* **2006**, *128*, 4970.

(18) Widawsky, J. R.; Kamenetska, M.; Klare, J.; Nuckolls, C.; Steigerwald, M. L.; Hybertsen, M. S.; Venkataraman, L. *Nanotechnology* **2009**, *20*, 434009.

(19) Lu, Q.; Liu, K.; Zhang, H.; Du, Z.; Wang, X.; Wang, F. *ACS Nano* **2009**, *3*, 3861.

(20) Hybertsen, M. S.; Venkataraman, L.; Klare, J. E.; Cwhalley, A.; Steigerwald, M. L.; Nuckolls, C. *J. Phys.: Condens. Matter* **2008**, *20*.

(21) Kamenetska, M.; Koentopp, M.; Whalley, A. C.; Park, Y. S.; Steigerwald, M. L.; Nuckolls, C.; Hybertsen, M. S.; Venkataraman, L. *Phys. Rev. Lett.* **2009**, *102*.

(22) Park, Y. S.; Whalley, A. C.; Kamenetska, M.; Steigerwald, M. L.; Hybertsen, M. S.; Nuckolls, C.; Venkataraman, L. *J. Am. Chem. Soc.* **2007**, *129*, 15768.

(23) Venkataraman, L.; Park, Y. S.; Whalley, A. C.; Nuckolls, C.; Hybertsen, M. S.; Steigerwald, M. L. *Nano Lett.* **2007**, *7*, 502.

(24) Quinn, J. R.; Foss, F. W.; Venkataraman, L.; Hybertsen, M. S.; Breslow, R. *J. Am. Chem. Soc.* **2007**, *129*, 6714.

(25) Quinn, J. R.; Foss, F. W.; Venkataraman, L.; Breslow, R. *J. Am. Chem. Soc.* **2007**, *129*, 12376.

(26) Breslow, R.; Foss, F. W. *J. Phys.: Condens. Matter* **2008**, *20*.

(27) Breslow, R.; Murayama, D. R.; Murahash.Si; Grubbs, R. *J. Am. Chem. Soc.* **1973**, *95*, 6688.

(28) Manatt, S. L.; Roberts, J. D. *J. Org. Chem.* **1959**, *24*, 1336.

(29) Roberts, J. D. *Chem. Soc., Spec. Publ.* **1958**, *12*, 111.

(30) Giese, B. *Annu. Rev. Biochem.* **2002**, *71*, 51.

(31) Gray, H. B.; Winkler, J. R. *Chem. Phys. Lett.* **2009**, *483*, 1.

(32) Jeuken, L. J. C.; Jones, A. K.; Chapman, S. K.; Cecchini, G.; Armstrong, F. A. *J. Am. Chem. Soc.* **2002**, *124*, 5702.

(33) Meyer, T. J.; Huynh, M. H. V.; Thorp, H. H. *Angew. Chem., Int. Ed.* **2007**, *46*, 5284.

(34) Yavin, E.; Boal, A. K.; Stemp, E. D. A.; Boon, E. M.; Livingston, A. L.; O'Shea, V. L.; David, S. S.; Barton, J. K. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 3546.

(35) Wu, W. P.; Liu, Y. Q.; Zhu, D. B. *Chem. Soc. Rev.* **2010**, *39*, 1489.

(36) Solomon, G. C.; Herrmann, C.; Vura-Weis, J.; Wasielewski, M. R.; Ratner, M. A. *J. Am. Chem. Soc.*, *132*, 7887.

(37) Solomon, G. C.; Vura-Weis, J.; Herrmann, C.; Wasielewski, M. R.; Ratner, M. A. *J. Phys. Chem. B*, *114*, 14735.

(38) Genereux, J. C.; Barton, J. K. *Chem. Rev.*, *110*, 1642.

(39) Seferos, D. S.; Blum, A. S.; Kushmerick, J. G.; Bazan, G. C. *J. Am. Chem. Soc.* **2006**, *128*, 11260.

(40) Seferos, D. S.; Trammell, S. A.; Bazan, G. C.; Kushmerick, J. G. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 8821.

(41) Martin, S.; Grace, I.; Bryce, M. R.; Wang, C. S.; Jitchati, R.; Batsanov, A. S.; Higgins, S. J.; Lambert, C. J.; Nichols, R. J. *J. Am. Chem. Soc.* **2010**, *132*, 9157.

(42) Schneebeli, S. T.; Kamenetska, M.; Cheng, Z.; Skouta, R.; Friesner, R. A.; Venkataraman, L.; Breslow, R. *J. Am. Chem. Soc.* **2011**, *133*, 2136.

(43) Winstead, B.; Ravaioli, U. *IEEE Trans. Electron Devices* **2000**, *47*, 1241.

(44) Larkem, A.; Larkem, H. *Indian J. Chem., Sect. B: Org. Chem. Incl. Med. Chem.* **2002**, *41*, 175.

(45) Park, Y. S.; Widawsky, J. R.; Kamenetska, M.; Steigerwald, M. L.; Hybertsen, M. S.; Nuckolls, C.; Venkataraman, L. *J. Am. Chem. Soc.* **2009**, *131*, 10820.

(46) Haenel, M. W.; Lintner, B.; Benn, R.; Rufinska, A.; Schroth, G. *Chem. Ber.* **1985**, *118*, 4922.

(47) Gundermann, K. D.; Roeker, K. D. *Justus Liebigs Ann. Chem.* **1976**, 140.

(48) Kwong, C.-Y.; Leung, M.-k.; Lin, S.-C.; Chan, T.-L.; Chow, H.-F. *Tetrahedron Lett.* **1996**, *37*, 5913.

(49) Haenel, M. W. *Tetrahedron Lett.* **1976**, 3121.

(50) Klapars, A.; Buchwald, S. L. *J. Am. Chem. Soc.* **2002**, *124*, 14844.

(51) Schneebeli, S.; Kamenetska, M.; Foss, F.; Vazquez, H.; Skouta, R.; Hybertsen, M.; Venkataraman, L.; Breslow, R. *Org. Lett.* **2010**, *12*, 4114.

(52) Eckertmaksic, M.; Fabian, W. M. F.; Janoschek, R.; Maksic, Z. B. *Theochem: J. Mol. Struct.* **1995**, *338*, 1.

(53) Wang, D. Z.; Streitwieser, A. *Theor. Chem. Acc.* **1999**, *102*, 78.

(54) Blatchly, J.; Taylor, R. *J. Chem. Soc. B.: Phys. Org.* **1968**, 1402.

(55) Field, L. D.; Sternhell, S.; Wilton, H. V. *J. Chem. Educ.* **1999**, *76*, 1246.

(56) Venkataraman, L.; Klare, J. E.; Tam, I. W.; Nuckolls, C.; Hybertsen, M. S.; Steigerwald, M. L. *Nano Lett.* **2006**, *6*, 458.

(57) McConnell, H. *J. Chem. Phys.* **1961**, *35*, 508.

(58) Nitzan, A. *Annu. Rev. Phys. Chem.* **2001**, *52*, 681.

(59) Fawcett, J. K.; Trotter, J. *Acta Cryst.* **1966**, *20*, 87.

(60) Breslow, R.; Schneebeli, S. *submitted*.

(61) Vollhardt, P.; Siegel, J. *Chem. Eng. news* **1996**, *April 1*.

(62) Gerkin, R. E.; Lundstedt, A. P.; Reppart, W. J. *Acta Cryst. C.: Cryst. Struct. Comm.* **1984**, *40*, 1892.

(63) Fu, J. M.; Snieckus, V. *Tetrahedron Lett.* **1990**, *31*, 1665.

(64) Palameta, B.; Zambeli, N. *J. Org. Chem.* **1964**, *29*, 1031.

(65) Cookson, R. C.; Parsons, P. J. *J. Chem. Soc.: Chem. Comm.* **1978**, 822.

(66) Cheng, Z. L.; Skouta, R.; Vazquez, H.; Widawsky, J. R.; Schneebeli, S.; Chen, W.; Hybertsen, M. S.; Breslow, R.; Venkataraman, L. *Nat. Nanotechnol.* **2011**, *6*, 353.

(67) Meisner, J. S.; Kamenetska, M.; Krikorian, M.; Steigerwald, M. L.; Venkataraman, L.; Nuckolls, C. *Nano Lett.* **2011**, *11*, 1575.

(68) Munakata, M.; Wu, L. P.; Ning, G. L.; Kuroda-Sowa, T.; Maekawa, M.; Suenaga, Y.; Maeno, N. *J. Am. Chem. Soc.* **1999**, *121*, 4968.

(69) George, S. J.; Ajayaghosh, A. *Chem. Eur. J.* **2005**, *11*, 3217.

(70) Higuchi, H.; Tani, K.; Otsubo, T.; Sakata, Y.; Misumi, S. *Bull. Chem. Soc. Jpn.* **1987**, *60*, 4027.

(71) Kannen T.; Sakata, Y. M., S., N.; Otsubo *Bull. Chem. Soc. Jpn.* **1976**, *49*, 3307.

(72) Otsubo, T.; Horita, H.; Misumi, S. *Synth. Commun.* **1976**, *6*, 591.

(73) Kamenetska, M.; Koentopp, M.; Whalley, A. C.; Park, Y. S.; Steigerwald, M. L.; Nuckolls, C.; Hybertsen, M. S.; Venkataraman, L. *Phys. Rev. Lett.* **2009**, *102*.

(74) Parameswaran, R.; Widawsky, J. R.; Vazquez, H.; Park, Y. S.; Boardman, B. M.; Nuckolls, C.; Steigerwald, M. L.; Hybertsen, M. S.; Venkataraman, L. *J. Phys. Chem. Lett.* **2010**, *1*, 2114.

(75) Schulz, J.; Vogtle, F. In *Cyclophanes*; Springer-Verlag Berlin: Berlin 33, 1994; Vol. 172, p 41.

(76) Shukla, M. K.; Dubey, M.; Leszczynski, J. *ACS Nano* **2008**, *2*, 227.

(77) Shukla, M. K.; Dubey, M.; Leszczynski, J. *Acs Nano* **2008**, *2*, 227.

(78) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(79) Winberg, H. E.; Fawcett, F. S. *J. Am. Chem. Soc.* **1960**, *82*, 1428.

(80) Longone, D. T.; Chow, H. S. *J. Am. Chem. Soc.* **1964**, *86*, 3898.

(81) Otsubo, T.; Mizogami, S.; Sakata, Y.; Misumi, S. *Tetrahedron Lett.* **1971**, *50*, 483.

(82) Hartner, H. (Merck Patent Gesellschaft mit beschränkter Haftung, USA). Process for the preparation of [2,2]-paracyclophane. US Patent 4,532,369, 1985.